Titre: Title:	Dynamic time slots allocation in WDM-TDM hybrid networks
Auteur: Author:	Hua Jiang
Date:	2005
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Jiang, H. (2005). Dynamic time slots allocation in WDM-TDM hybrid networks [Master's thesis, École Polytechnique de Montréal]. PolyPublie. https://publications.polymtl.ca/7633/

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/7633/
Directeurs de recherche: Advisors:	
Programme: Program:	Unspecified

UNIVERSITÉ DE MONTRÉAL

DYNAMIC TIME SLOTS ALLOCATION IN WDM-TDM HYBRID NETWORKS

Hua JIANG
DÉPARTEMENT DE GÉNIE ÉLECTRIQUE
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION DU DIPLÔME DE
MAÎTRISE ÈS SCIENCES APPLIQUÉES
JUIN 2005

Your file Votre référence
ISBN: 978-0-494-16800-4
Our file Notre référence
ISBN: 978-0-494-16800-4

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé:

DYNAMIC TIME SLOTS
ALLOCATION IN WDM-TDM HYBRID NETWORKS

présenté par: **Hua JIANG**
en vue de l'obtention du diplôme de: Maîtrise ès science appliquées
a été dûment acceptée par le jury d'examen constitué de:

Jean-François FRIGON, Ph.D., président
Burnilde SANSÒ, Ph.D., directrice de recherche
André GIRARD, Ph.D., membre

Acknowledgments

I would like to take this opportunity to thank many people for their great help and support through the course of the studies.

I am very grateful to Professor Brunilde Sansò, my research director, for her constant help and continuous guidance. She corrected my directions and gave me suggestions every time when I encountered difficulties. I thank her as well for providing me the financial support during my study.

I would like to thank to Professors André Girard and Jean-François Frigon for giving me very useful advice as members of the jury. I also need to thank to all my colleagues in LORLAB for their encouragement and friendship. The Résumé and the Condensé en Français in this thesis are translated by them.

Finally, I will express my deepest appreciation to my husband, my baby and my parents for being the loving source throughout the years.

Résumé

Les réseaux optiques peuvent répondre au besoin croissant en bande passante dans le systèmes, puisque WDM et OTDM offrent deux techniques possibles pour exploiter la bande passante des fibres optiques. Récemment, la méthode de multiplexage WDM-TDM a été proposeé pour améliorer l'utilisation des ressource en bande passante. Un réseau hybride WDM-TDM utilise la méthode de multiplexage WDM-TDM.

Dans ce travail, nous nous intéressons au problème d'allocation de bande passante dans les réseaux hybrides WDM-TDM. Plus précisement, nous abordons le problème d'affectation dynamique des intervalles de temps à une connexion. Ce problème consiste à déterminer le nombre d'intrevalles de temps nécessaires qu'il faut affecter à une connexion pour garantir ses éxigence en Qualité de Service (QdS). Nous voulons aussi mettre à jour périodiquement l'affectation pour suivre la variation du trafic. L'objectif est d'accroître l'efficacité d'utilisation des intervalles de temps tout en satisfaisant la Qualité de Service(QdS).

Un méthode simple mais efficace a été proposée dans ce travail. L'affectation des intervalles de temps est basée sur l'information historique telle que la quantité d'intervalles de temps affectés et la performance du système au cours des périodes précédentes. Des simulations ont été réalisées pour évaluer la performance de la méthode proposée. Cette performance est comparée à celle des méthods conventionnelles d'allocation. Les résutats numériques ont montré que la méthode proposée est efficace et que la performance du système a été améliorée par rapport aux méthodes conventionnelles d'affectation d'intervalles de temps.

Abstract

The optical network can provide the answer to the acute need for the very-high-bandwidth transmission systems, since WDM and OTDM offer two possible techniques to exploit the bandwidth of optical fibers. More recently, the WDM-TDM multiplexing method has been proposed to make the most of the bandwidth resource. A WDM-TDM hybrid network employs the WDM-TDM multiplexing method.

In this project, we deal with the bandwidth allocation problem in WDM-TDM hybrid networks. We concentrate on the time slot assignment to a connection in a real-time manner. The problem that we confront is how many time slot are needed to assign to a connection in order to guarantee its QoS requirement. Moreover, we would like to update the assignment periodically to follow the varying traffic. The goal is to increase the time slot utilization efficiency while satisfying the QoS.

A time slots allocation method is proposed in this project. The time slots are assigned based on the history information such as the number of time slots assigned and the system performances in the previous periods. Simulations have been done in order to evaluate the performance of the method proposed. The results show that the proposed method worked effectively and the system performances have been improved comparing with the conventional time slots allocation methods.

Condensé en Français

Introduction

Une méthode hybride, appelée multiplexage WDM-TDM, a été récemment proposée pour optimiser l'utilisation de la ressource de la bande passante. Cette méthode consiste à d'abord diviser la bande passante selon les longueurs d'ondes, et ensuite de diviser chacune des longueurs d'ondes dans le temps. Les réseaux hybrides WDM-TDM utilisent donc le multiplexage WDM-TDM. L'objectif de ce projet est de trouver une méthode efficace pour réaliser l'allocation dynamique des intervalles de temps (time slots) dans une seule connexion.

Énoncé du problème

Dans les réseaux hybrides WDM-TDM, des connexions avec différents types de service se déplacent de leurs origines vers leurs destinations en suivant des chemins optiques préétablis. Chaque connexion occupera un certain nombre d'intervalles de temps pré-assignés. Chaque type de service requiert ses propres critères de Qualité de Service (QdS). Le trafic est composé d'un flot de paquets de tailles variables. Puisque le taux d'arrivée des paquets d'une connexion n'est pas constant et varie tout le temps, l'allocation pour cette connexion devrait être mise à jour de temps en temps pour s'adapter au trafic. Cependant, nous assumons que l'allocation sera maintenue durant une certaine durée de temps, appelée *période*.Les intervalles de temps ne seront réallouées qu'à la fin de chaque période, Une période contient un cetain nombre de trames TDM, et ces dernières contiennent un ensemble d'intervalles de temps. La QdS à respecter durant l'assignation est le taux de perte de paquets. Le problème est donc énoncé de la façon suivante: Sachant que, pour une connexion, nous connaissons le nombre d'intervalles de temps assignées ainsi que le taux de pertes de paquets résultant dans chacune des périodes précédentes, il faut déterminer le nombre d'intervalles de temps qui nous sont nécessaires.

Système de file d'attente

Nous supposons une file d'attente par connexion. Les paquets arrivant vers la file d'attente sont ceux de la connexion en question. Les intervalles de temps réservés jouent le rôle de serveur de file d'attente. Les paquets sont de longueur variable, et leurs arrivées suivent une distribution générale. Il existe un nombre très limité de places dans la mémoire tampon, ainsi qu'expliqué auparavant. La discipline de service sera de faire la tâche la plus courte en premier. En résumé, la file d'attente dans ce projet est de type G/G/1/m.

Techniques d'empaquetage

Les paquets arrivants sont d'abord insérés dans la file d'attente et sont ensuite empaquetés dans les intervalles de temps réservés, suivant l'approche connue sous le nom de *méthode 1-bin* . Nous supposons que le nombre d'intervalles de temps assignés sera réservé continuellement. Les intervalles de temps réservés seront considérées comme un seul bin. Cette technique d'empaquetage est illustrée à la figure 1. Les trois intervalles de temps réservés forment un seul bin, et les paquets sont empaquetés dans ces intervalles. Si la partie non remplie n'est pas suffisante pour accommoder le prochain paquet, ce dernier doit rester dans la file d'attente et attendre la prochaine trame TDM. Le routeur n'a pas besoin d'identifier chacuns des intervalles de temps. Au lieu de cela, il extraira seulement les données entre deux bits de synchronisation.

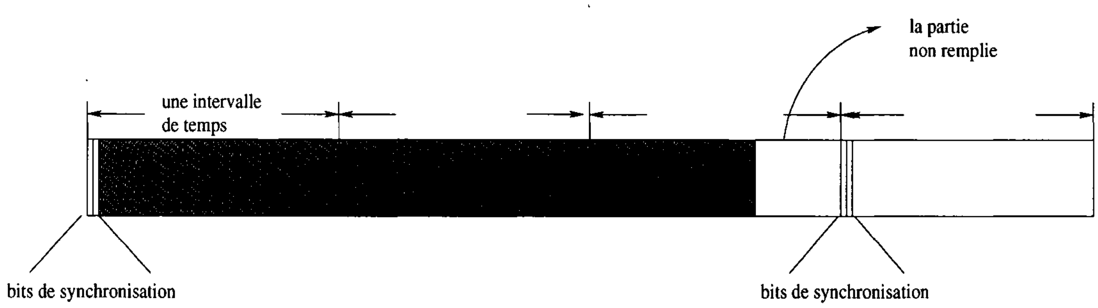

Figure 1: La technique d'empaquetage 1-bin

L'algorithme d'allocation

Nous définissons qu'un assignement est le nombre d'intervalles de temps qui sont réservés dans chaque trame TDM d'une période, pour une connexion. Une assignation pondéré est obtenue en mettant des poids sur les assignations de plusieurs périodes précédentes. Une assignation compensée est une assignation utilisée pour compenser des assignations utilisant des poids.

L'algorithme d'allocation est décrit dans la figure 2. À la fin de chaque période, une certaine quantité d'information sera récupérée. En fonction de cette information, une assignation pondéré et le taux de perte de paquets correspondant seront calculés pour la période suivante. Grâce à l'estimation du taux de perte de paquets, un ajustement de l'assignation pondéré sera implémenté si le critère d'ajustement n'est pas respecté. Autrement, l'assignation pondéré courante sera gardée. De cette façon, nous obtenons le nombre d'intervalles de temps à assigner à la prochaine période.

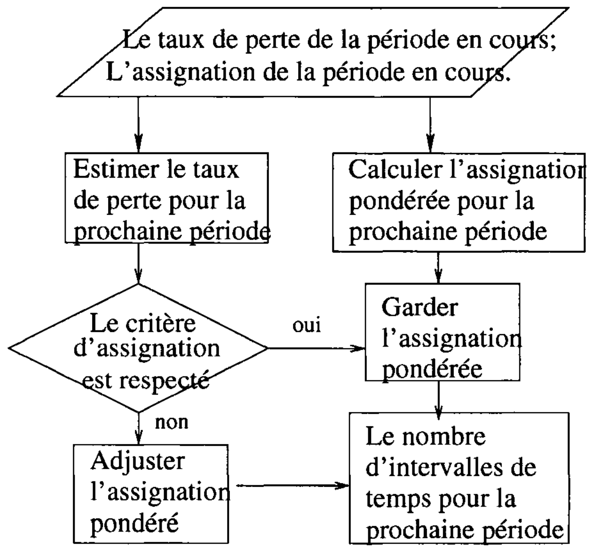

Figure 2: L'algorithme D'allocation

Calculer de l'assignation pondéré et estimer le taux de perte

Nous allons explorer deux approches pour calculer l'assignation pondérée et estimer le taux de perte pour une prochaine période. La 1^{re} approche est décrite par les formules suivantes:

$$W_{i+1} = N_i \tag{1}$$

$$E_{i+1} = \alpha_1 R_i + \alpha_2 R_{i-1} + ...\alpha_n R_{i-n+1} \tag{2}$$

$$\sum_j \alpha_j = 1 \qquad j = 1, ...n \tag{3}$$

W_i: l'assignation pondéré de la période i.

N_i: l'assignation de la période i.

E_i: le taux de perte de paquet estimé de la période i.

R_i: le taux de perte de paquet réel de la période i.

α_i: le poids associé à la période i.

La $2^{ème}$ approche est décrite par les formules suivantes.

$$W_{i+1} = \alpha_1 N_i + \alpha_2 N_{i-1} + ...\alpha_n N_{i-n+1} \tag{4}$$

$$E_{i+1} = \alpha_1 R_i + \alpha_2 R_{i-1} + ...\alpha_n R_{i-n+1} \tag{5}$$

$$\sum_j \alpha_j = 1 \qquad j = 1, ...n \tag{6}$$

6 cas avec des poids différents sont considérés dans ce projet. Tous les 6 cas seront explorés en utilisant les deux approches.

Le critère et la formule d'adjustment

Soit R_t le *seuil de taux perte de paquets*, qui peut être considéré comme la QdS associée au trafic. Un *intervalle acceptable* $[R_t^-,\ R_t^+]$ est une fenêtre centrée sur le seuil de perte. Quand le taux de perte estimé appartient à l'intervalle acceptable, un ajustement de l'allocation des intervalles de temps n'est pas nécessaire. D'autre part, si le taux de perte estimé n'appartient pas à l'intervalle acceptable, un ajustement doit être fait. O_{i+1} est assignation compensée et est définie par l'expression.

$$O_{i+1} = \left\lfloor \left(c\frac{(E_{i+1} - R_t)\bar{A}L_p}{L_s M} \right) + 0.5 \right\rfloor \tag{7}$$

$\bar{A}$ (paquets): le nombre moyen de paquets qui arrivent pendant une période.

L_p (octets): la taille moyen d'un paquet.

L_s (octets): longueur d'un intervalle.

M: le nombre de trames TDM dans une période.

c: coefficient d'ajustement.

Le coefficient c est introduit pour obtenir un entier supérieur à 1 comme décrit dans la figure 3. La raison derrière ces trois définitions est que nous voulons un système qui réagit différement face à des situations différentes.

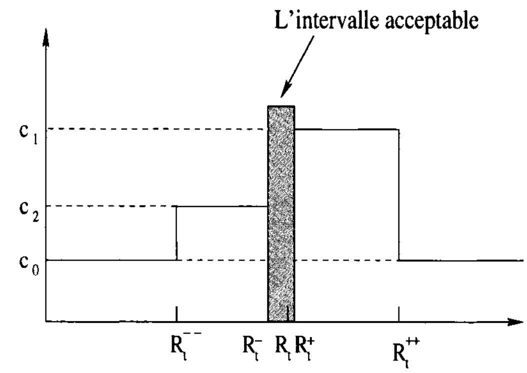

Figure 3: La définition de c.

L'équation d'évaluation de performance

La performance d'un algorithme d'allocation peut être évaluée par des mesures différentes. Nous choisissons de faire une mesure pondérée comme suit.

$$Z = P_0\varepsilon - P_1\eta - P_2V \tag{8}$$

P_0 : Le bénéfice gagné par augmentation de l'efficacité d'un intervalle de temps de 1%

P_1 : Le coût à payer quand on augmente le taux d'occupation de la bande passante de 1%.

P_2 : Le coût à payer quand on augmente l'écart type de la demande en QdS de 1%

V : L'écart type du taux de perte par rapport au seuil du taux de perte.

ε: le taux d'utilisation des intervalles de temps, qui devrait être le plus haut possible.

η: Le taux d'occupation de la bande passante qui devrait être le plus bas possible.

Quand $\bar{R} \leq R_t$ est satisfaite, plus Z est grand, plus le système est performant. Parcontre, quand cette contrainte n'est pas satisfaite, Z peut aussi prendre une grande valeur. Une fonction pénalité Γ est définie en but d'avoir une équation d'évaluation de performance Z universelle.

$$\Gamma = \begin{cases} 0, & \text{si } \bar{R} \leq R_t, \\ \gamma, & \text{autrement} \end{cases} \tag{9}$$

Où γ est une valeur grande, comme par exemple 200. En suite, l'équation d'évaluation des performances peut être réécrite comme il suit. P_3 indique le coût à payer pour la violation des exigences de QdS.

$$Z = P_0\epsilon - P_1\eta - P_2V - P_3\Gamma \tag{10}$$

La Modèle de simulation

Des données réelles prises sur Internet sont utilisées comme trafic en entrée dans la simulation de ce projet. Quelque modifications ont été faites pour rendre le trafic plus adapte à les réseaux optiques.

Le modèle de simulation est montrée dans la figure 4. Un certain nombre d'intervalles de temps de chaque trame TDM est réservé pour le trafic d'une connexion. Les paquets en entrée, de longueur différente, seront introduits dans une file d'attente si elle n'est pas pleine. Dans le cas où elle est pleine, le paquet sera perdu. Quand une trame TDM arrive, les paquets dans la file d'attente seront transmis sur les intervalles de temps réservés. Les paquets qui n'ont pas pu être transmis attendront dans la file d'attente pour une autre trame TDM. L'algorithme qui implémente le modèle est décrit dans la figure 5. Tous les programmes de la simulation de ce projet sont écrits en Java. La simulation est en partie basée sur SSJ.

Résultat des simulations

Les simulations principales traitent tous les 6 cas des deux approches. Les résultats montrent que la perte moyenne des paquets dans tous les cas satisfait les exigences

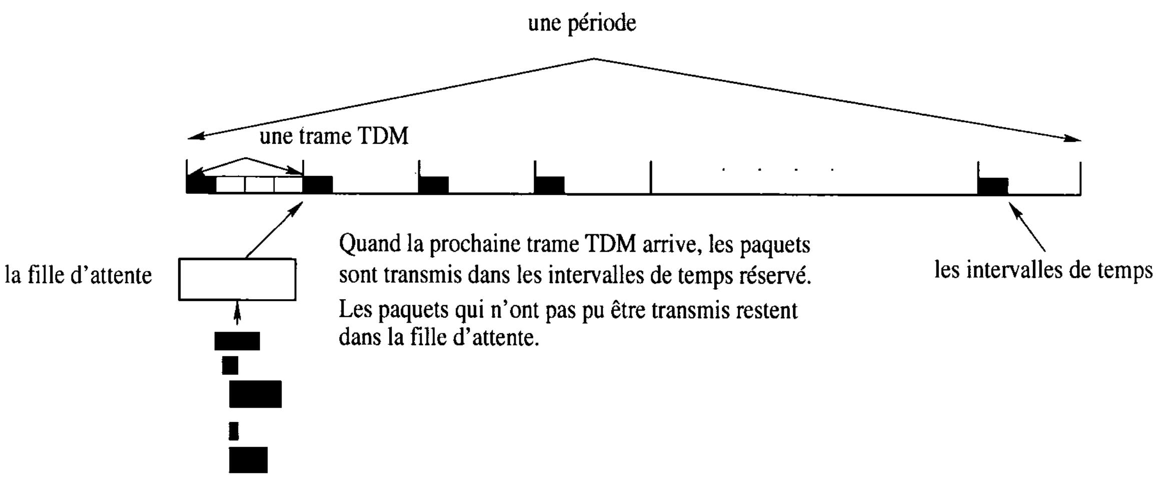

Figure 4: Modèle de simulation

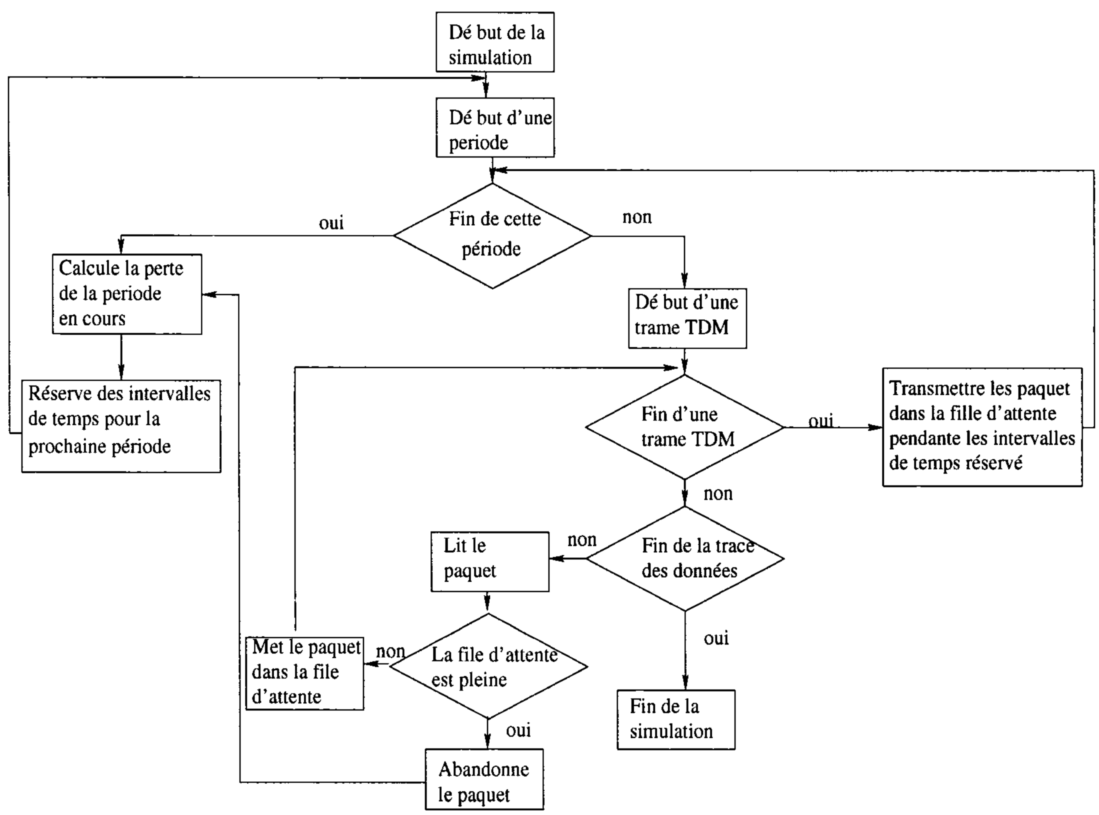

Figure 5: Algorithme de simulation

de Qualité de Service (QdS). Et les cas de la $2^{ème}$ approche travaillent mieux que ceux de la 1^{ere}. Notre méthode, comparée à une allocation fixe de 7 intervalles de temps et à l'allocation à la fréquence crête de 22 intervalles de temps, a des performances meilleures. Cependant, nous avons aussi noté que pendant les simulations principales, le système a eu un comportement pauvre autour de la 300^{ieme} periode dans la majorité des cas, et ce comportement diminue beaucoup les performances. Pour résoudre ce problème, nous avons considéré de fixer deux seuils, supérieur et inférieur, pour l'assignation de chaque période, et de garantir que l'assignation se passe toujours en respectant les seuils. En suite, d'autres simulations ont été faites (avec des *Bollinger Bands* et des *Moving Average Envelopes*).

Les résultats montrent que l'augmentation du taux de perte du paquet autour de la 300^{eme} période ne se produit plus et que l'on a une nette amélioration de la valeur qui évalue la performance. Nous avons aussi trouvé que les MAEs fonctionnent mieux que les BBs puisqu'il pourrait y avoir une convergence avec les BBs.

Enfin, il serait intéressant de tester cette méthode avec d'autres sources, car le traceur de données utilisé dans ce projet présente peut-être des particularités. De plus, les performances avec des connexions multiples sont également intéressantes. Cependant, on remarquera que la technique du 1-bin pourrait nécessiter quelques changements dans le système TDM, et plus de recherches devraient être faites sur ce sujet.

Contents

List of Tables

List of Figures

List of Abbreviations

WRN: Wavelength Routing Network.

RWA: Routing and Wavelength Assignment.

RWTA: Routing, Wavelength and Time Slot Assignment.

WDM: Wavelength Division Multiplexing.

TDM: Time Division Multiplexing.

OTDM: Optical Time Division Multiplexing.

QoS: Quality of Service.

BBs: Bollinger Bands.

MAEs: Moving Average Envelopes.

SSJ: Stochastic Simulation in Java.

List of Definitions

A TDM Frame: A frame consists of a set of time slots. In this thesis, a TDM frame is defined containing 200 time slot.

Period: A portion of time during which the time slot assignment does not change and at the end of which the assignment will be updated. In this thesis, a period is defined as containing 5000 TDM frames.

An assignment: An assignment is the number of time slots that are reserved in every TDM frame in a period for a connection.

A weighted assignment: An assignment obtained by weighting the assignments of several previous periods.

An offset assignment: An assignment to compensate the weighted assignment.

Time slots utilization efficiency: The amount of bits effectively transmitted divided by the amount of bits that could have been transmitted by the time slot that are assigned during the transmission of a connection.

Bandwidth occupation ratio: The total number of time slot assigned to a connection divided by the total number of the available time slot during the whole session of this connection.

Acceptable range: An acceptable range is a range centered with the loss threshold.

List of notations

$\bar{A}$: the mean amount of packets that arrive in a period.

A_i: the number of packets that arrive in period i of the modified real data trace.

A_i^o: the number of packets that arrive in period i of the original real data trace.

B_d : the number of bytes transmitted successfully during the transmission session of a connection.

B_t : the number of bytes that could be transmitted by the slots that were assigned.

c: adjustment coefficient.

D_i : the i^{th} data value.

E_i: the estimated packet loss rate for period i.

F: the transmission speed of a wavelength (Gbps).

i, j: the period counter.

k: the number of periods that a connection lasts.

L_i^B : the i^{th} lower bound value.

L_p: the mean packet length (bytes).

L_s: slot length (bytes).

m : the number of data being averaged when calculating the moving average. A typical value is 20.

M: the number of TDM frames in a period.

$\bar{M}_i$: the i^{th} moving average value.

n: the number of previous periods that we take into account for the weighted assignment.

n_s: the sample number when calculating the confidence interval.

N_i: the assignment of period i.

O_i: the offset assignment of period i.

P_0 : the benefit we will gain from increasing the time slot utilization efficiency by one percent.

P_1 : the expense we should pay for augmenting the bandwidth occupation ratio by one percent.

P_2 : the expense we should pay for augmenting the standard deviation from QoS requirement by one percent.

P_3 : the expense we should pay for violating the QoS requirement.

$\bar{R}$: the average packet loss rate.

R_i: the packet loss rate of period i.

R_t: the threshold of packet loss rate.

R_t^+: the upper bound of the acceptable range.

R_t^-: the lower bound of the acceptable range.

R_t^{++}, R_t^{--}: define the bounds when the estimated loss rate is much higher/lower than the threshold.

S_f : the number of time slots in a TDM frame.

U_i^B : the i^{th} upper bound value.

V : the standard deviation of the loss rate from loss rate threshold.

W_i: the weighted assignment of period i.

Z: performance evaluation value.

α_i: the weight associated with period i.

γ: the penalty value defined in the penalty function.

Γ: the penalty function used in the performance evaluation equation.

Δ: standard error used in the calculation of the confidence interval. ε: the time slot utilization efficiency.

η: the bandwidth occupation ratio.

μ_i: the amount of packets that are going to be added or deleted to period i of the original real data trace.

Chapter 1

Introduction

1.1 WDM-TDM Hybrid Networks

The recent years have witnessed an outstanding growth in the area of telecommunication and computer networking, leading to an acute need for very-high-bandwidth transmission systems. Optical fibers can supply the required bandwidth. *Wavelength Division Multiplexing* (WDM) offers one possible technique to exploit the bandwidth of optical fibers. In WDM, the optical transmission spectrum of a fiber is divided into a number of non-overlapping wavelength bands. Another technique is to use *Optical Time Division Multiplexing* (OTDM), in which a fiber link is considered as a single channel and is divided into time slots. More recently, *the WDM-TDM multiplexing* has been proposed to make the most of the bandwidth resource, in which the bandwidth resource is first wavelength divided and then each wavelength is further time divided. A WDM-TDM hybrid network employs the WDM-TDM multiplexing method.

1.2 Motivation and Objective

How to make the most of the resource in order to increase the network throughoutput is a crucial aspect in network planning. In the literature, the resource allocation problem is usually defined as follows: given a matrix of bandwidth requirement connections, assign the bandwidth resource to the connections in a way to satisfy the bandwidth requirement matrix. The element in the matrix represents the bandwidth requirement of a single connection. The amount of bandwidth requirement of a connection is normally fixed and is determined based on the long-term statistical data or the peak arrival rate. It is obvious that these methods will either lead to a poor QoS or cause severe resource waste. How to determine the bandwidth requirement of a single connection more efficiently is the motivation of this project.

In this project, we deal with the bandwidth allocation problem in WDM-TDM hybrid network. We concentrate on the time slot assignment to a connection in a real-time manner. The problem that we confront is how many time slot are needed to assign to a connection in order to guarantee its QoS requirement. Moreover, we would like to update the assignment periodically to follow the varying traffic. The goal is to make the most of the resource, specifically it means to increase the time slot utilization efficiency and decrease the bandwidth occupation ratio, while satisfying the QoS.

The objective of the project is to find an effective method to realize the dynamic time slot allocation. We determine the allocation based on the history information such as the number of the time slots assigned and the system performances in the previous periods. We would like to see how this method will work compared with the conventional methods.

Most research work about the bandwidth allocation or admission control is based on the assumption of a certain traffic model. The Poisson process is the traffic model most commonly used in the telecommunication area. However, it is known that the Poisson model fails as the accurate packet arrival process for most wide area network arrival processes [25, 5]. More recently, more complicated models [19] have been proposed to simulate the Internet traffic. Nevertheless, the network traffic is composed of complex random processes. Consequently, it might not be tractable by the theoretical models. In our project, we prefer to neglect the theoretical model and to adopt a real data trace as the input traffic.

We assume that the traffic in question is of certain type of real time application. In order to strictly guarantee the delay, a small buffer capacity is provided in the router along the path from the origin to the destination. The QoS that should be respected during the assignment is the packet loss rate. Our objective is then to assign the bandwidth to the traffic stream while satisfying its QoS.

1.3 Thesis Organization

In Chapter 2, the main concepts of optical networking will be first reviewed, then the WDM-TDM hybrid network will be presented. The problem of interest and the solution approaches are defined in detail in Chapter 3. Chapter 4 talks about the modifications of the real data trace. The simulation model and parameters are listed in Chapter 5. The main simulation and the results are presented and analyzed in Chapter 6. Extension simulation results with Bollinger Bands and Moving Average Envelopes are discussed in Chapter 7. The conclusions are drawn in chapter 8.

Chapter 2

Literature Review

2.1 Multiplexing Methods in Optical Networks

Wavelength Division Multiplexing (WDM) [17, 22, 2] is the multiplexing technique most widely used in optical networks. In WDM, the optical transmission spectrum of a fiber is divided into a number of non-overlapping wavelength bands, see figure 2.1 (b). For instance, a fiber link with spectrum 10 Gbps can be divided into 100 to 1000 discrete wavelengths. Later, we will see that in wavelength routed optical networks, a wavelength is used as the degree of freedom to route the optical signal while in optical packet switched networks, a wavelength simply serves as a channel.

Another technique is called *Optical Time Division Multiplexing* (OTDM) [9, 20, 3], in which a fiber link is considered as a single wavelength and is divided into time slot, see figure 2.1 (a).

More recently WDM-TDM multiplexing method [18, 27, 21, 12] was proposed to make the most of the bandwidth resource, in which the bandwidth resource is first wavelength divided and then each wavelength is further time divided, see figure 2.1 (c).

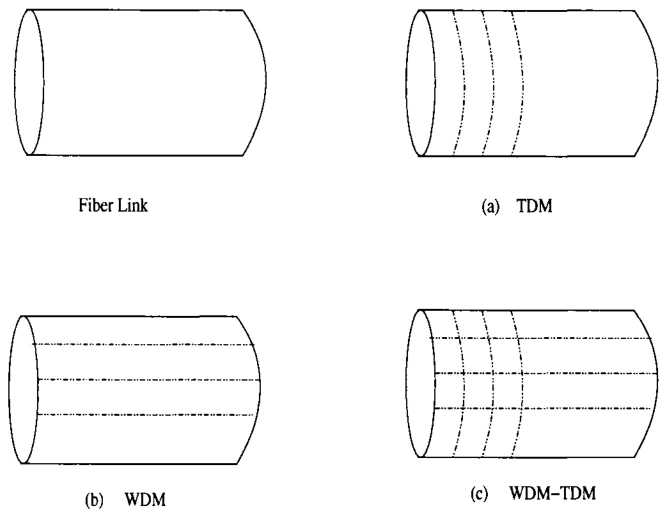

Figure 2.1: Three multiplexing techniques

2.2 Evolution of Optical Networks

2.2.1 Wavelength Routed Network

The WDM based wavelength routed network is the promising network structure for the next generation of wide-area backbone networks. A typical wavelength routed network is shown in figure 2.2. The network consists of routing nodes interconnected by wavelength-division-multiplexed (WDM) fiber links to form an arbitrary physical topology. End-users are connected to a routing node via an access node. Each fiber link is divided into several wavelengths and each wavelength can be switched and routed independently.

In a wavelength routed network, a wavelength will be exclusively reserved by a traffic flow, and once it is reserved no other traffic flows can share this wavelength any more. Accordingly, a traffic flow can not traverse the network until a path is established. This path is called a *lightpath* [3]. A lightpath is an optical communication channel between two routing nodes in the network which might span more than

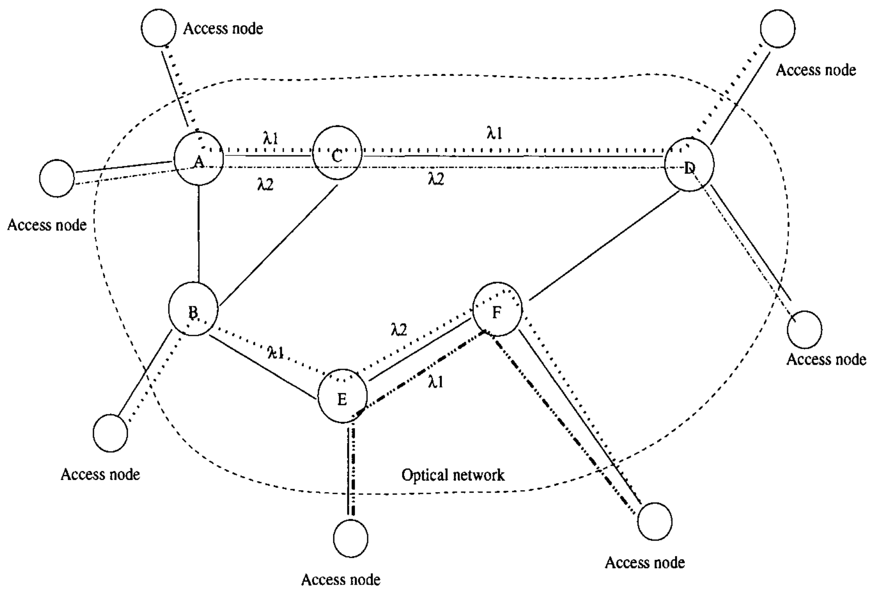

Figure 2.2: Wavelength routed network

one fiber link. The intermediate nodes in the fiber path route the lightpath in the optical domain. No optical/electrical transformation is needed.

Before a traffic flow can enter into the wavelength routed network, a lightpath should be established. During the transmission period, the traffic flow will exclusively occupy the lightpath. After the transmission is finished, the lightpath is released and the wavelengths along the lightpath become free again. Once a lightpath is established, the routing node can route the traffic flow according to its coming node and wavelength instead of referring to the header of the packets.

In the absence of any wavelength conversion device, a lightpath is required to be on the same wavelength throughout its path in the network; this requirement is referred to as *the wavelength continuity constraint.* For example, in figure 2.2, routing node C does not have conversion capability and the wavelength λ_1 along the lightpaths $A - C - D$ should remain the same. This requirement may not be necessary if we also have wavelength converters in the network such as the routing node E. Another requirement is that two or more lightpaths traversing the same fiber link must be on different wavelength channels so that they do not interfere with one another. This is called *wavelength distinction constraint.* For instance, in figure 2.2, there are two lightpaths traveling through fiber link $E - F$. Then the wavelengths of the two lightpaths should be distinct.

2.2.2 Hybrid Switched Networks

The problem with wavelength routed networks is that the bandwidth utilization efficiency is comparatively low. Recall that a traffic flow will unconditionally occupy the whole wavelength during its transmission, no matter if it has continuous data or not. When wavelength conversion is available, this mechanism works exactly as the traditional circuit switched network. The industry then have tried to figure out a way to implement fine granularity packet switching in the optical domain just like the way it works in the Internet, in order to make good use of the bandwidth resource. This results in the optical packet switched network.

In optical packet switched networks, a wavelength is not used by a single traffic flow, it simply serves as an independent bandwidth resource available to any packets from any source to any destination. The packet is routed independently according to the information found in its header and will be routed into any free bandwidth resource.

A typical switch node in optical packet switched network is shown in figure 2.3 [26]. Packets are aligned in the input interface. The header of the packet is extracted by the interface and afterward it is sent into the control unit, where the optical signal is transformed into an electrical one and a switching decision is made according to the information in the header. The packet is then switched into an output port in a wavelength. In the output interface the transformed optical signal is added to the packet. Obviously, in this process, optical/electrical transformation is needed, which will decrease the transmission speed.

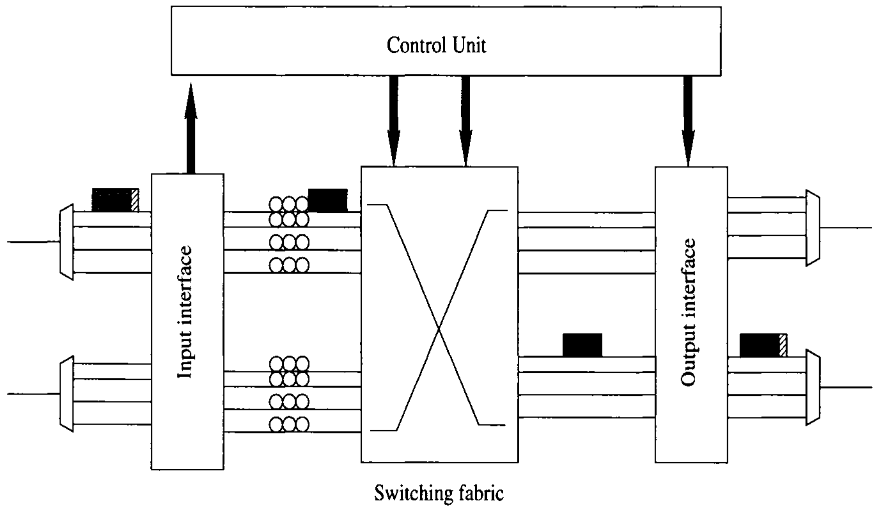

Figure 2.3: A typical optical packet switching node cited from [26]

Another difficulty that the optical packet switched network confronts is contention resolution, as contention occurs very often in packet switched networks. In traditional packet switched networks, contention is resolved by store-and-forward techniques. However, because of the lack of optical RAM, we have to resort to other methods. Fiber delay lines are one of the solution [4]. Packets in contention are delayed in fiber delay lines and are sent out later. However, the fiber lines are bulky and costly as well. For typical optical network, readers can refer to [14] and [11].

Wavelength routed networks utilize the bandwidth resource poorly while optical packet switched network is somewhat immature. Therefore, the *WDM-TDM hybrid network* technology has recently been explored.

2.2.3 WDM-TDM Hybrid Networks

As mentioned above, WDM is a well developed method to use the optical bandwidth. However, sometimes a wavelength is too large for certain traffic requirements. For instance, for bursty traffic, a whole wavelength may constitute a waste of resources as it may require a fractional wavelength. Thus, there is a need to segment the wavelength into fractions. Provisioning fractional wavelength is achieved by dividing a wavelength into time slot and multiplexing traffic on the wavelength, which results in the *WDM-TDM hybrid networks*[18, 27, 21, 12].

Figure 2.4 shows a WDM-TDM optical network. Each fiber contains w wavelength. The bandwidth of a wavelength is further partitioned into fixed-length time slot. A TDM frame consists of a fixed number of time slots. If the router is equipped with a time slot interchanger, it will then be able to route the traffic from one time slot to another. For each wavelength, the routing node behaves as a traditional TDM circuit switching node. If the router is equipped with wavelength converter as well, it will be able to route the traffic from one time slot in a wavelength to another time slot in another wavelength.

The optical path of a connection is thereby no longer a simple lightpath [12]. In the case without wavelength converter, the lightpath is composed of the preassigned time slot in the TDM frame of the wavelengths along the predetermined path in the network. In the case with wavelength converter, the lightpath will include not only the time slot information but also the preassigned wavelengths along the path. After the optical path is constructed, the route will transmit packets on the assigned time slot of the assigned wavelength(s).

In the example in figure 2.4, two lightpaths are established, 1 and 2. The node B router does not have wavelength converter capability, the time slot are changed from time slot 1 and 2 in fiber link $A-B$ to time slot 2 and 3 in fiber link $B-C$ but all data remain on λ_1. Node D, on the other hand, does have wavelength converter capability and the traffic is routed from time slot 1 and 3 in wavelength λ_1 to time slot 1 and 2 but this time on a different wavelength λ_2.

Constructing a WDM-TDM hybrid network requires routers capable of wavelength conversion as well as time slot interchange. An experimental photonic time slot interchanger is tested in [23]. Two kinds of router structures equipped with wavelength conversion and time slot interchange are proposed and tested in [12, 18].

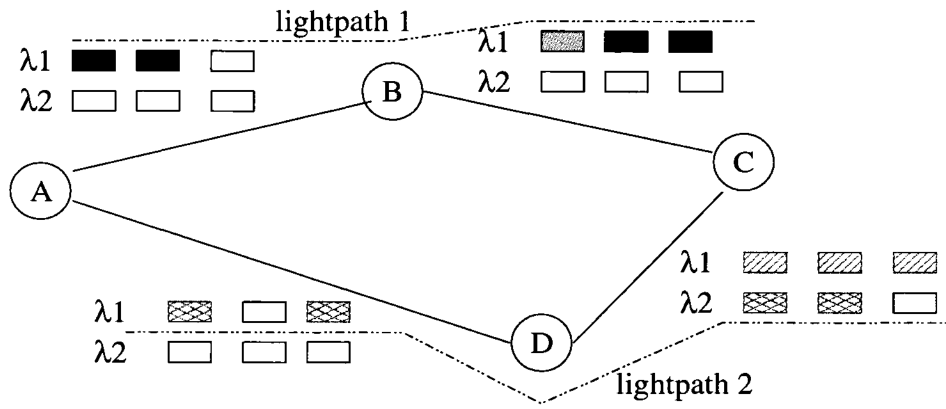

Figure 2.4: WDM-TDM optical network

2.3 Bandwidth Allocation Problem

2.3.1 Bandwidth Allocation Problem in Optical Networks

The bandwidth allocation is a key problem existing in all kinds of networks. In wavelength routed networks, the bandwidth allocation problem is also called routing and wavelength assignment (RWA). The problem is often stated as follows: given a set of lightpath requirements, normally represented by a matrix, route and assign a wavelength to each of them. A connection sometimes is considered to occupy a single lightpath [28]. Or, in other cases, the lightpath requirements matrix is given without considering the relationship between a lightpath and a connection [3].

As explained in section 1.1, in a WDM-TDM network, a lightpath includes not only the wavelength information but also the time slot information. The bandwidth allocation problem is now the routing, wavelength and time slot assignment problem (RWTA). In [12], the authors consider the RWTA problem as determining the lightpaths for a given set of connection requests with bandwidth demand. Moreover, the bandwidth demand of each connection is regarded as fixed and it is previously obtained by long-term statistics. Fixed allocation to a connection according to its long-term behavior will inevitably cause poor QoS or severe bandwidth waste. To determine the bandwidth requirement for a connection or a traffic stream between an origin-destination pair is exactly the problem considered here.

2.3.2 Modeling Bandwidth Allocation as a Bin Packing Problem

Resource management problems in telecommunications can sometimes be modeled as bin packing problems and solved by one of the available algorithms. For instance, bin packing is used to solve the channel/slots allocation problem in GPRS [24], in TDMA Satellite System [16] and in Time-Division duplex CDMA system [13].

Bin packing [7] refers to the problem of finding the minimum number of bins of length capacity W into which item with length $w_1, ...w_n$, $w_i \leq W$ can be packed. If we regard a time slot as a bin and the packets as the items to be packed, the bandwidth allocation problem can be considered as a bin packing problem.

If the item weights are of certain stochastic distribution, then the bin packing problem are specifically called as *stochastic bin packing problem* [15, 10, 6, 8]. In the literature, there are two types of stochastic bin packing: bin packing with stochastic item length and bin packing with stochastic item arrival pattern.

The bin packing with stochastic item length is defined as follows: given a set of independent random variables $S = \{X_1, X_2, ..., X_n\}$ and an overflow probability p, partition the set S into the smallest number of sets $S_1, S_2, ...S_k$ such that

$$Pr\Big[\sum_{i:X_i \in S_j} X_i > 1\Big] \leq p, \quad \forall 1 \leq j \leq k.$$

The basic idea is to try to find a way as accurate as possible to approximate the probabilistic item size by deterministic one. Therefore, the stochastic bin packing can be transformed into deterministic bin packing, so that the well-developed bin packing algorithms can be applied to solve the original stochastic problem.

[15] discussed the problem with on-off sources, while [10] discussed the problem with Poisson source and exponential source.

In bin packing with stochastic arrival patterns, the problem with fixed item sizes and certain type of stochastic arrival process is typically treated. [6] studied the problem with Poisson arrivals and [8] studied the problem with more general arrival processes. The analysis aims at the stability of the queue length.

It is known that the bin packing problem is NP-hard. The situation is much more complicated in the stochastic context. We can see from the above discussion that in either stochastic case a regular distribution of the source is assumed first. However, recall that in this project, a real data trace will be employed as the input traffic, which means that we will have neither a regular packet length distribution nor a regular packet arrival pattern. Therefore, the models discussed in literature do not apply to our case. We need to propose another approach to deal with the time slot allocation problem.

Chapter 3

Problem Statement and Solution Approaches

3.1 The Problem Statement

The problem being investigated in this project is defined as follows. In WDM-TDM hybrid networks, connections of different service types travel from origins to destinations along the pre-established lightpaths. Each connection will occupy some pre-assigned time slots. Each service type has its own QoS requirements. We will only consider the time slot assignment problem occurring in the ingressing routing node.

The traffic is a stream of packets of various lengths. Since the packet arrival rate of a connection is not constant, the allocation for this connection should be updated from time to time so as to adapt to the traffic. However, we assume that the allocation will be kept intact during a certain time, called *period*. The time slots will only be reallocated at the end of each period. A period contains a number of TDM frames, which in turn contains a set of time slot. An amount of time slots of each TDM frame are reserved for a connection.

In this project, the traffic is assumed to follow a real time application. In order to strictly guarantee the delay, a very small buffer capacity is provided in the router. The QoS that should be respected during the assignment is then the packet loss rate.

Given that, for a connection, we know the number of assigned time slots and the resulting packet loss rate in each of the previous periods, the problem that we face is to determine the time slot needed in the next period trying to satisfy the QoS requirements. In brief, the problem can be expressed as (see figure 3.1):

In period i, given:

the QoS requirement of a connection,

the number of time slots assigned to each of the previous periods,

and the measured packet loss rate of the each of the previous periods,

Find:

How many time slots are needed in period $i + 1$ for this connection?

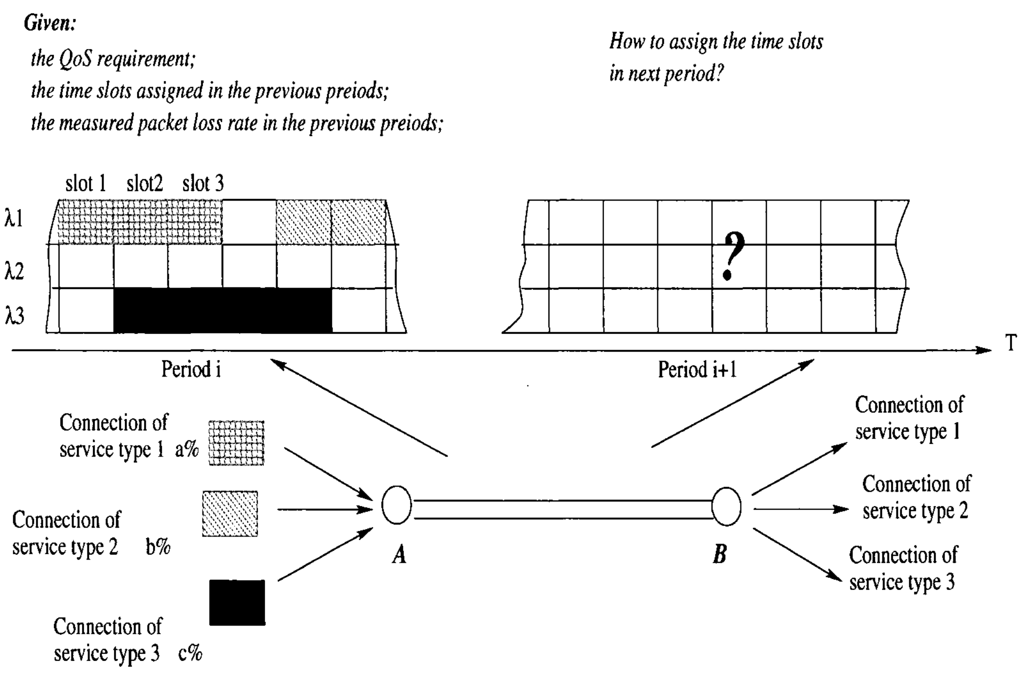

Figure 3.1: The definition of the problem

We propose to solve this problem in the following manner. At the end of a period, a weighted assignment for the next period will be calculated based on the historical assignments and system performance (details will be discussed in section 3.5). We will then determine, according to certain method, whether the weighted assignment will fit the coming packets in the next period or not. If the answer is positive the weighted assignment will be kept into the next period, otherwise some adjustments will be implemented.

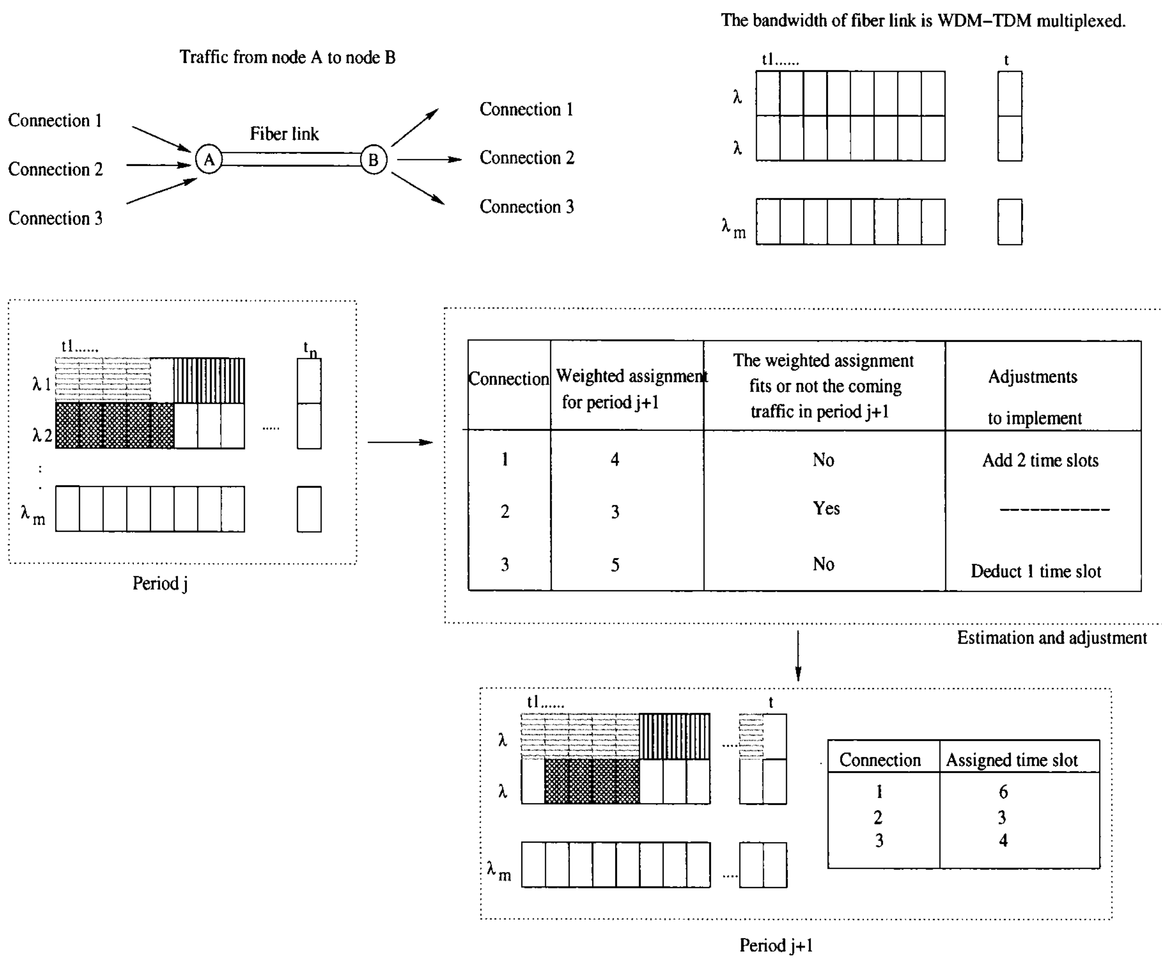

Figure 3.2: An illustrative example

A simple illustrative example is shown in figure 3.2. Traffic of three connections flow from node A to node B, the fiber link A-B is wavelength and time divided. 4 time slots, 3 time slots and 5 time slots are assigned to connection 1, 2 and 3 respectively in period j. Based on previous calculations, the weighted assignments to the period $j+1$ are all the same. It is estimated that for connection 1 and 3 the weighted assignment will not fit the coming packets in period $j+1$ while it will for connection 2. Therefore adjustments must be invoked to connection 1 and 3 while the weighted assignment for connection 2 is kept.

During the process, we focus on how to calculate the weighted assignment, how to estimate whether it will fit the demand in the next period or not and on how to implement the adjustments.

Since the way of assigning the time slots to different connections is identical, only one connection is considered in this project.

3.2 Queuing System

As mentioned in section 3.1, the router is equipped with a buffer of small capacity. The features of this queuing system is described in this section.

Since only one connection is considered in this project, it is quite natural that one queue per connection is assumed. As shown in figure 3.3, the arrivals at the queue are the packets of the connection in question. The reserved time slots work as the server. Assume that F is the transmission speed of a wavelength, S_f the number of time slots in a TDM frame and N is the number of assigned time slots to the connection, the service rate for this connection will equal to NF/S_f. The service rate will change from period to period, since the assignment will be updated from period to period.

The packets are of various lengths which means that we have a general service process. Furthermore, the arrival of the packets are of general distribution. There is

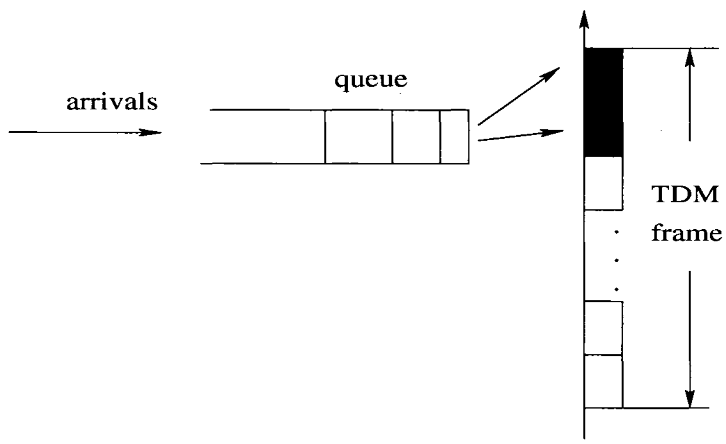

Figure 3.3: The queuing system

very limited number of places in the buffer, as explained before. The packets have to wait in the queue before being served. The arriving packets will be dropped if the queue is full, which contributes to the packet loss rate. In order to decrease the mean waiting time and the packet loss rate, the service discipline will be the shortest job first. As one queue per connection is assumed, we do not have sharing issue in this system. To sum up, the queuing system in this project could classified as a G/G/1/m system.

3.3 Packing Techniques

The incoming packets will be first inserted into the queue and will be packed into the reserved time slots later on. And as mentioned before, the packets can be of different lengths. On the one hand, several small packets can be packed into one time slot. On the other hand, it is very likely that the length of a packet is longer than that of a time slot. In that case, a big packet will occupy several time slots. How to pack the packets into the time slots will be explained in detail in this section.

There are two possible packing techniques, which will be presented hereafter. Nevertheless, recall that the service discipline in this queuing system is the shortest job first, which means that the packets are going to be sorted in length-ascending order before being packed into the reserved time slots.

3.3.1 The 2-arrays Method

The first possible method is called *2-arrays method.* The sorted packets will be divided into two arrays. The packets whose lengths are shorter than the length of a time slot are arranged in the first array. The remaining packets are arranged in the second array. In both arrays, packets are listed in length-ascending order.

The packets in the first array will be packed into the reserved time slots first, as shown in figure 3.4 (a). If the unfilled part of a time slot is not enough to accommodate the next packet, this part will be left unfilled and a new time slot will be used, if available, to carry the next packet.

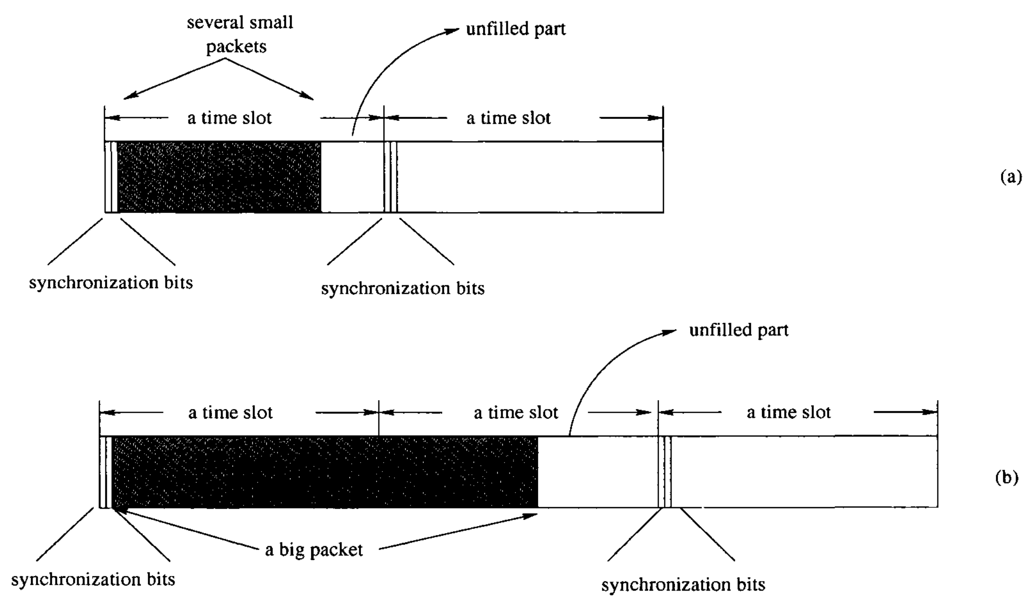

Figure 3.4: The 2-arrays packing technique

When all the packets in the first array are packed and if there are still free time slots available, packets in the second array will be packed as shown in figure 3.4 (b). If all the reserved time slots in the current TDM frame are used up, the remaining packets should wait in the queue for the next TDM frame.

In figure 3.4, the synchronization bits in both cases will help the router to identify the arrival of a new time slot. Once the router detect a new time slot is coming, the data between two synchronization bits will be read.

3.3.2 The 1-bin Method

The second possible technique is called *1-bin method.* Since only one connection is considered, it is assumed in this method that the number of time slots assigned to this connection will be reserved continuously instead of discretely by the router, see figure 3.5. The assumption is reasonable in the sense that, from the point of view of the administrator, it is trouble-saving to try the best to assign several continuous time slots to a connection instead of discrete ones.

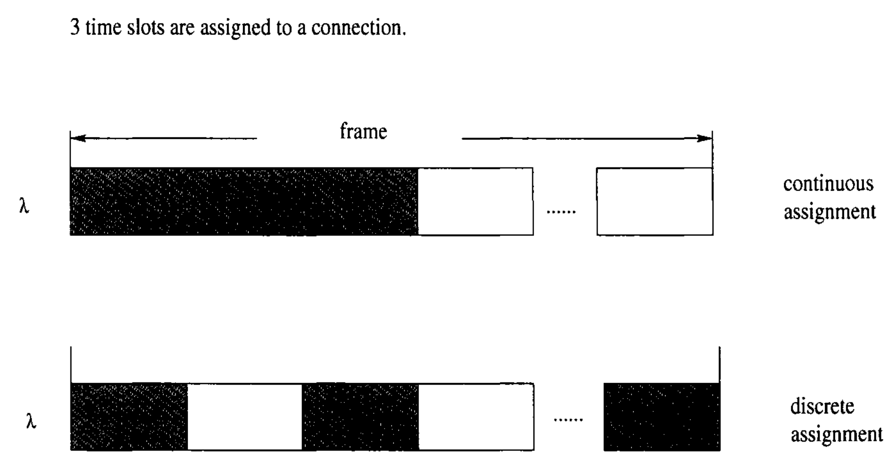

Figure 3.5: The continuous assignment vs. discrete assignment

With the above assumption, the reserved time slots will be regarded as one bin in this method. The sorted packets are the items to be packed. This packing technique is illustrated in figure 3.6. In this example, the reserved 3 time slots are looked as one bin, and the sorted packets are packed into these time slots. If the unfilled part is not enough to accommodate the next packet, the packet has to stay in the queue and wait for the next TDM frame. The router does not need to identify every time slot. It will only extract the data between two synchronization bits, no matter if there is one time slot or several continuous slots.

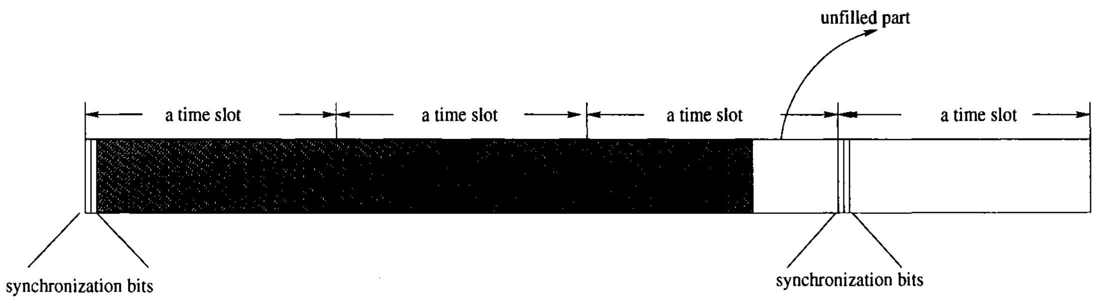

Figure 3.6: The 1-bin packing technique

In both methods, the unfilled part of a time slot contributes to the bandwidth waste. It is clear that the 1-bin method will have less bandwidth waste. Moreover, the packing algorithm is simpler comparing with the one of the first method. Therefore, the 1-bin method is the one that we have decided to use in this project.

It is worth mentioning that in order to accommodate long packets, we have assumed the TDM system a little bit different from the classical one. To be specific, the synchronization bits are used differently. The reasons behind this are explained hereafter. In practice, it is very often that the packet lengths exceed the capacity of a time slot. There are two ways to deal with it. Either cut the long packet into pieces or leave the packets as they are and pack them anyway. The first method will increase the complexity to the upper layer, as packet identification and reorganization are needed. The complexity is added to the network or link layer in the second method. In this project, we choose to leave the packets as they are, which then requires some change of the TDM system.

One of the proposals to implement the 1-bin method is the following. Once the lightpath is established for a connection, the reservation information will be propagated from the origin to the destination. By doing this, a routing node will know which time slots are reserved for this connection. For example, time slot 1, 2 and 3 are reserved. It is then possible to control the synchronization-bits-generator to generate synchronization bits only to slot 1 and 4, while skip slot 2 and 3. In a word, in order to implement the packing technique to reality, more work has to be done, and it is also one of our future works.

3.4 The Allocation Algorithm

Now the proposed allocation algorithm will be presented. As discussed in section 3.1, at the end of each period, the time slots allocation will be updated so as to adapt to the changing traffic. The objective of the allocation algorithm is to assign as few time slots as possible to a connection while meeting the QoS of this connection.

Before discussing the allocation algorithm, three important definitions are first introduced. An assignment is the number of time slots that are reserved in every TDM frame in a period for a connection. A weighted assignment is obtained by weighting the assignments of several previous periods. An offset assignment is an assignment to compensate the weighted assignment.

The allocation algorithm is described in figure 3.7. At the end of each period, some information, such as the packet loss rate and the assignment of the current period, will be collected. Based on these data, a weighted assignment for the next period will be calculated and the corresponding packets loss rate will be estimated according to a method to be described. With the estimation of the packet loss rate, an adjustment to the weighted allocation will be implemented if the adjustment criteria is not respected. Otherwise, the weighted assignment will be kept. In this

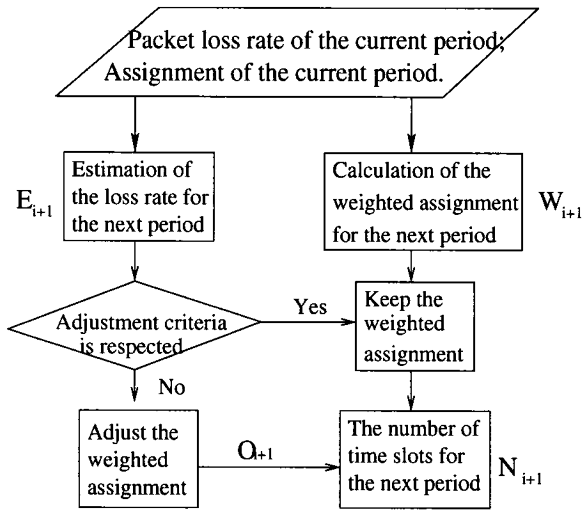

Figure 3.7: The allocation algorithm

way, we obtain the number of time slots to be assigned to the next period. Each step in the allocation algorithm will be explained in detail in the rest of this chapter.

3.5 Weighted Assignment Calculation and Loss Rate Estimation

As illustrated in the algorithm shown in figure 3.7, the offset assignment will be added to the weighted one or not depending on the loss rate estimation for the next period. The weighted assignment calculation and the loss rate estimation will be discussed in this section.

The calculation of the weighted assignment and the estimation of the packet loss rate are done based on the historic data. We decide to do it in a weighted way. We are interested in finding out how many previous periods should be traced back and how these periods should be taken into account. To answer the first question, the parameter n representing the number of the former periods taken into account is defined. To answer the second question, we associate certain weights with the periods under consideration.

In this project, we will explore two approaches as follows. In both cases, we are in period i and try to allocate the time slots for the period $i+1$.

3.5.1 The packet loss rate weighted approach

In this approach, only the packet loss rate for the next period is weighted by those of the previous periods. The weighted assignment will be the amount of time slots assigned to the current period. Therefore, we will have:

$$W_{i+1} = N_i \tag{3.1}$$

$$E_{i+1} = \alpha_1 R_i + \alpha_2 R_{i-1} + ... \alpha_n R_{i-n+1} \tag{3.2}$$

$$\sum_j \alpha_j = 1 \qquad j = 1, ...n \tag{3.3}$$

W_i: the weighted assignment of period i.

N_i: the assignment of period i.

E_i: the estimated packet loss rate for period i.

R_i: the actual packet loss rate of period i.

α_i: the weight associated with period i.

3.5.2 The packet loss rate and assignment weighted approach

In this approach, both the weighted assignment and the packets loss rate are identically weighted by the previous periods. We will then have:

$$W_{i+1} = \alpha_1 N_i + \alpha_2 N_{i-1} + ... \alpha_n N_{i-n+1} \tag{3.4}$$

$$E_{i+1} = \alpha_1 R_i + \alpha_2 R_{i-1} + ... \alpha_n R_{i-n+1} \tag{3.5}$$

$$\sum_j \alpha_j = 1 \qquad j = 1, ...n \tag{3.6}$$

In these two approaches, we have defined the weighted assignment calculation function and the packet loss rate estimation function. In practical, the estimated value will be compared with the QoS requirement to determine the weighted assignment should be kept or should be adjusted. Different values of n and α will be considered so as to find the best value.

3.6 Adjustment Criteria

A *packet loss rate threshold* is defined in this project. The goal is to try to keep the packet loss rate within the threshold. The packet loss rate threshold can be considered as the QoS associated with the traffic. In what follows, we will use the words threshold and QoS equivalently.

Let us define now an *acceptable range* as a window centered on the loss threshold. For example, if the loss threshold is defined as $R_t = 2\%$, the acceptable range could be [1.2% - 2.3%].

The adjustment criteria is defined based on the acceptable range. When the estimated loss rate falls inside the acceptable range, we will say that the QoS will be satisfied in the next period. Then the criteria is said to be respected and the time slots allocation adjustment is not necessary. If, on the other hand, the estimated loss rate falls outside the range, there should be an adjustment. For instance, with the acceptable range defined as [1.2% – 2.3%], if E_{i+1} equals to 2.23%, there is no need to adjust the weighted assignment. On the other hand, if E_{i+1} equals to 2.7%, an allocation adjustment will then be needed. Let R_t^+ and R_t^- represent respectively the upper and lower bound of the range. We have $R_t^+ = 2.3\%$ and $R_t^- = 1.2\%$

The values of the acceptable range are chosen by the network administrator. Since the upper bound defines to what degree the estimated loss rate will be accepted, the closer to the threshold the stricter the system is. The choice of the lower bound is looser as we are more tolerable to the situation when the loss rate is lower than the threshold. In that case, an adjustment will increase the utilization of the network.

3.7 Adjustment Formula

If the estimated packet loss rate of the next period falls into the acceptable range, the weighted assignment for the next period is left unchanged. Otherwise, the allocation adjustment is changed. In mathematical terms, this can be explained as follows.

$$N_{i+1} = \begin{cases} W_{i+1} & \text{if } R_t^- \le E_{i+1} \le R_t^+ \\ W_{i+1} + O_{i+1} & \text{otherwise.} \end{cases} \tag{3.7}$$

O_{i+1} is the offset assignment, which could be positive or negative. The question is how many time slots should be added or removed when an adjustment needs to be implemented? The basic idea is as follows. In the case when more time slots are needed so as to satisfy the QoS requirement, we will first calculate how many packets would be dropped according to the estimated value of the packet loss rate and the threshold. We will then try to figure out how many time slots would be needed to carry those packets.

We define O_{i+1} as:

$$O_{i+1} = \left\lfloor \left(c \frac{(E_{i+1} - R_t)\bar{A}L_p}{L_s M} \right) + 0.5 \right\rfloor \tag{3.8}$$

where

$\bar{A}$ (packets): the mean amount of packets that arrive in a period.

L_p (bytes): the mean packet length.

L_s (bytes): slot length.

M: the number of TDM frames in a period.

c: adjustment coefficient.

The intuition behind the formula is as follows. $(E_{i+1} - R_t)$ represents the difference between the estimated loss rate and the threshold. When the estimated loss rate is higher than the threshold, $E_{i+1} - R_t$ is positive. The loss rate difference multiplied by the mean amount of packets that arrive in a period $(E_{i+1} - R_t)\bar{A}$ gives the amount of packets that are estimated to be dropped in the next period but we should be transmitting in order to satisfy the QoS. The result further multiplied by the mean packet length L_p will then provide the result in term of bytes. Finally $(E_{i+1} - R_t)\bar{A}L_p/(L_s M)$ will be the amount of time slots per frame needed to be able to transmit those bytes in the next period.

In the same way, when $E_{i+1} < R_t$ which means less time slots are required by the period $i+1$, $(E_{i+1} - R_t)\bar{A}L_p/(L_s M)$ will represent the amount of the excessive time slots. With a negative sigh, these excessive time slots will be reduced.

However, $\frac{(E_{i+1}-R_t)\bar{A}L_p}{L_s M}$ will most of the time give us a very small fraction which will not result in any change in allocation. Therefore, c is introduced in order to get an integer larger than 1. For example, suppose that $\bar{A} = 10637$ packets, $L_p = 285$ bytes, $L_s = 200$ bytes, $M = 5000$, $R_t = 2\%$. If in a period the packet loss rate is 2.4%, $\frac{(E_{i+1}-R_t)\bar{A}L_p}{L_s M}$ results in 0.01213 which means only 0.01213 more time slot is needed. As a matter of fact, we can obtain that 1 more time slot is not needed until the packet loss rate goes higher than 34.99%. It is obviously not practical. With the introduction of $c = 100$, 1.213 (rounded to 1) slot is needed, which makes much more sense.

From the above discussion, it is obvious that the higher the value of c the more sensitive the system is. Three adjustment coefficients, c_0, c_1 and c_2, are defined as shown in figure 3.8. R_t is the loss rate threshold. $[R_t^-,\ R_t^+]$ defines the acceptable range. c_1 corresponds to the situation when the estimated loss rate is higher than R_t^+, c_2 corresponds to the opposite situation. c_0 corresponds to the case when the estimated loss rate is much higher or lower than the threshold, and in figure 3.8 R_t^{++} and R_t^{--} represent the upper and lower bound of this case. The reason for the three definitions is that we want the system to react differently to the different situations.

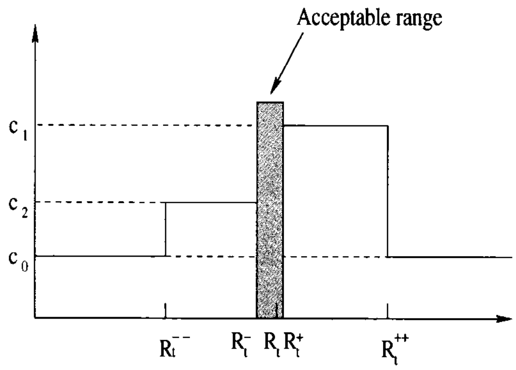

Figure 3.8: The definition of c.

Since we are more concerned with the case when the loss rate is higher than the threshold, we want the system to be more sensitive in this situation, and a big value is given to c_1. On the contrary, we are more tolerable when the loss rate is lower than the threshold, and a small value is given to c_2. The reason that we have c_0 is as follows. Occasionally, the amount of traffic might change suddenly and dramatically for a very short time and then go back to normal. However, it is preferable for the system not to respond to this occasional event instantly and rapidly. Therefore, a very small value will be given to c when the estimated loss rate is very high or very low.

We now explain how to define the adjustment coefficient c. The acceptable range of the packet loss rate will help us to define it. An example presented below will explain how it is going to work.

If we have a packet loss rate threshold of 2%, and the acceptable range is defined as [1.2%, 2.3%]. This means that, if the loss rate in period $i+1$ is estimated higher than 2.3% or lower than 1.2%, then the offset assignment is needed, i.e. at least 1 time slot should be added or removed. This can be written as:

$$O_{i+1} \begin{cases} \geq 1, & \text{if } E_{i+1} - R_t > 0.3\% \\ \leq -1, & \text{if } E_{i+1} - R_t < -0.8\% \end{cases} \tag{3.9}$$

Furthermore, in this example we suppose that $\bar{A} = 10637$ packets, $L_p = 285$ bytes, $L_s = 200$ bytes, $M = 5000$ TDM frames. Substitute these values into formula 3.8, we will have:

$$O_{i+1} = \begin{cases} \left\lfloor \Big(c_1(E_{i+1} - R_t) \times 10637 \times 285/(200 \times 5000) \Big) + 0.5 \right\rfloor, & \text{if } E_{i+1} > R_t \\ \\ \left\lfloor \Big(c_2(E_{i+1} - R_t) \times 10637 \times 285/(200 \times 5000) \Big) + 0.5 \right\rfloor, & \text{if } E_{i+1} < R_t \end{cases} \tag{3.10}$$

Combine (3.9) and (3.10), we can get $c_1 \approx 100$ and $c_2 \approx 41$. The definition of c_0 is less strict, and is defined as $c_0 = c_1/2$ in this project.

3.8 Performance Evaluation Measure

The allocation algorithm described in the previous sections aims to reduce the bandwidth waste while respecting the QoS of a connection. In order to evaluate how well the bandwidth is used, an evaluation formula is developed to measure the performance of this algorithm. In addition, by comparing the performance values of different allocation algorithms, we can also tell which algorithm is better than the others. Before introducing the equation, two important measurements are presented first.

3.8.1 Time Slot Utilization Efficiency

Recall that the *1-bin method* was chosen to pack the packets in the queue in the reserved time slots. As mentioned in 3.3.2, if the next packet in the queue is larger than the unfilled part of the reserved time slots (see figure 3.9) it has to stay in the queue and wait for the reserved time slots in the next TDM frame. The unfilled part of the time slots then contributes to the bandwidth waste. A better allocation method will result less waste and thus have higher utilization efficiency. The *time* *slots utilization efficiency* is then defined and used to indicate how well the *assigned* time slots are used, the higher the better.

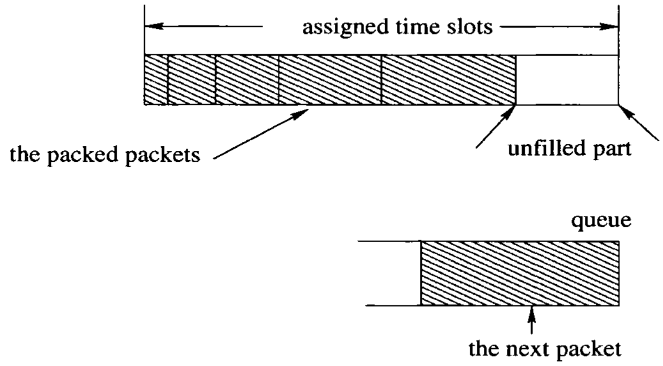

Figure 3.9: The wasted bandwidth.

Assume that B_d is the amount of bytes effectively transmitted, and B_t is the amount of the bytes that could have been transmitted by the slots that are assigned during the whole transmission session of a connection, as shown in figure 3.10. The time slots utilization efficiency of full ratio for allocated slots then equals to B_d divided by B_t. That is:

$$\varepsilon = \frac{B_d}{B_t} \tag{3.11}$$

We calculate B_t as follows. As shown in figure 3.10, given that N_i is the number of assigned time slots, L_s is the slot length in bytes and M is the number of TDM frame in a period, $N_i L_s M$ is the number of bytes that could have been transmitted in a period. Therefore, if k is the number of periods that a connection uses, we will have:

$$B_t = \sum_{i=1}^{k} N_i L_s M \tag{3.12}$$

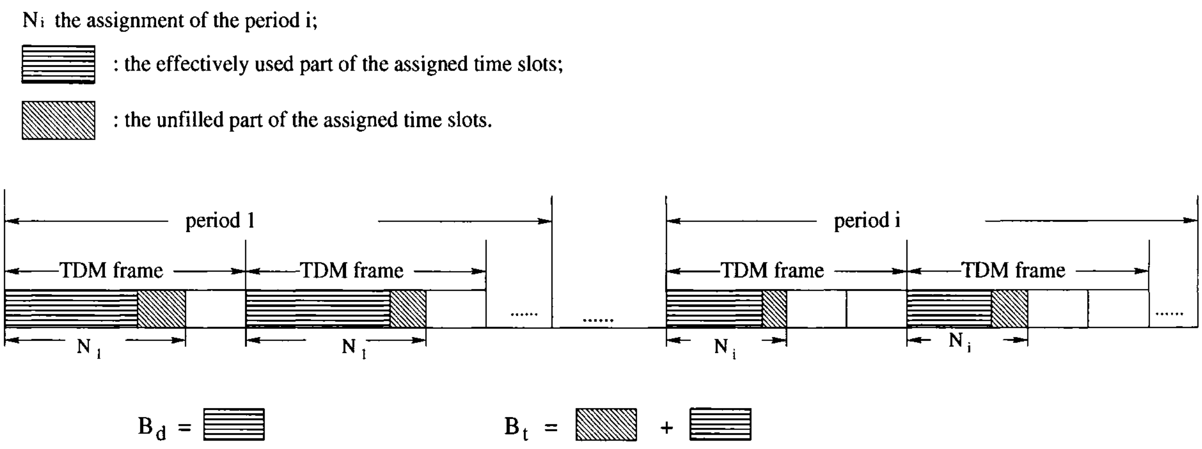

Figure 3.10: The time slot utilization efficiency.

3.8.2 Bandwidth Occupation Ratio

The *bandwidth occupation ratio* is defined as the bandwidth resource that are occupied by a connection. It equals to the time slots assigned to this connection divided by the total available time slots, as showed in formula 3.13.

$$\eta = \frac{\sum_{i=1}^{k} N_i M}{kMS_f} = \frac{\sum_{i=1}^{k} N_i}{kS_f} \tag{3.13}$$

An allocation method resulting in a low bandwidth occupation ratio while respecting the QoS is deemed to be better, as it allows the network accept more connections.

3.8.3 Performance Evaluation Measure

The performance of an allocation algorithm can be evaluated by several measurements. We want some of them as high as possible while others as low as possible. As we mentioned in previous sections, a high time slot utilization efficiency implies a good allocation method. On the other hand, a high bandwidth occupation ratio will indicate a bad one. In addition, an allocation resulting in a low standard deviation from the QoS threshold is desired as well. The reason is the following. Even with a low mean loss rate, chances are that the loss rate of each period varies a lot. However, these loss rate are preferred close to the QoS, which implies a low standard deviation from QoS. Finally, the QoS should be satisfied above all. In this project, we choose to do a weighted measurement of all the above. A performance evaluation measure Z is then defined as follows:

$$Z = P_0\varepsilon - P_1\eta - P_2V \tag{3.14}$$

P_0 : the benefit we will gain from increasing the time slot utilization efficiency by one unit.

P_1 : the expense we should pay for augmenting the bandwidth occupation ratio by one unit.

P_2 : the expense we should pay for augmenting the standard deviation from QoS requirement by one unit.

V : the standard deviation of the loss rate from loss rate threshold.

The first item in the evaluation equation represents the benefit we could gain from increasing the time slots utilization efficiency. The second and the third items represent the expense we have to pay for a high bandwidth occupation ratio and a high standard deviation of the packet loss rate.

In the project, we will compare the value of Z for different allocation methods. When the QoS requirement is satisfied, the following constraint is satisfied.

$$\bar{R} \leq R_t \tag{3.15}$$

The higher Z the better the system performs. However, when (3.15) is not satisfied, we might have a high value of Z as well. The reason is that the high packets loss rate is likely the consequence of insufficient of time slots allocation which indicates a low bandwidth occupation ratio η. Under this situation, it is then meaningless to compare Z.

In order to have a universal performance evaluation equation Z, we prefer to take the constraint (3.15) into account in the function Z and eliminate this constraint. In order to do so, a penalty function Γ is defined as follows:

$$\Gamma = \begin{cases} 0, & \text{if } \bar{R} \leq R_t, \\ \gamma, & \text{otherwise} \end{cases} \tag{3.16}$$

where γ is a large enough value, such as 200. Γ can be considered as a penalty for violating the QoS requirement. If the requirement is not satisfied, a penalty will be added to the evaluation equation. With the penalty Γ, the performance evaluation equation (3.14) can be rewritten as follows:

$$Z = P_0\epsilon - P_1\eta - P_2V - P_3\Gamma \tag{3.17}$$

The fourth item in (3.17) is the penalty function we add. P_3 is the expense we should pay for violating the QoS requirement.

In this way, when the QoS requirement is not satisfied we will have a very low or even negative value Z, which will indicate a poor performance.

Chapter 4

Traffic Source

4.1 Traffic Modifications

In our project, a method to allocate the time slots to connections has been proposed and a simulation will be implemented to examine the performance of the method. Instead of using classical distributions such as Poisson, we will test the method with a real trace obtained from the Internet. The reason for this is that, although classical distributions have been deeply studied and are therefore more understandable and controllable, they do not completely apply to real life, [25, 5]. One of the purposes of the project is to evaluate the method with more irregular and unpredictable traffic.

The real data trace used in this project was made by Digital Western Research Lab. The traces contain the wide-area traffic between Digital Equipment Cooperation and the rest of the world. The data trace file is a text file consisting of two columns. Every single row in this trace file represents a packet. One column represents the packet arrival time and the second column is the packet length.

This trace was collected in 1995. When we tried to implement our algorithm with this trace we found out that the traffic in the trace was not heavy enough for an optical network. Also, the amount of arrivals in each period did not vary a lot from a period to another. Consequently, slot allocation adjustments were not required very often. Therefore, some modifications were needed to make the traffic more suitable to the optical networks.

First, we needed to increase the arrival rate. In order to do so, the time duration of the trace is scaled from 3600 seconds down to 360 seconds. The arrival rate then increased 10 times.

Second, we needed to have different amounts of arrivals. we doubled the difference of the arrivals between two consecutive periods, i.e.

$$A_i - A_{i-1} = 2(A_i^o - A_{i-1}^o) \tag{4.1}$$

where

A_i: the number of packets that arrive in period i of the original data trace.

A_i^o:the number of packets that arrive in period i of the modified data trace.

We first calculated the amount of arrivals in each period of the original trace. Then, based on the formula (4.1) we obtained the expected amount of arrivals in each period of the modified trace. From this we obtained the amount of packets μ_i that should be added or deleted in period i.

These amounts of packets were added in the following way. Since the arrival pattern of the real trace is unknown and it probably will not be a regular pattern such as a Poisson distribution, the packets that are going to be added are then generated in a random way.

Notice that a row represents a packet and that the row number is the sequence number of the packet. In the case that the packets should be deleted, an amount of μ_i row numbers were generated by random number generator in Java and the corresponding packets were deleted from the trace file. In this way, the difference of the two consecutive periods is doubled.

Figure 4.1 shows the arrivals in each period of the trace before and after the modification 2. Be aware that the trace has already been scaled down.

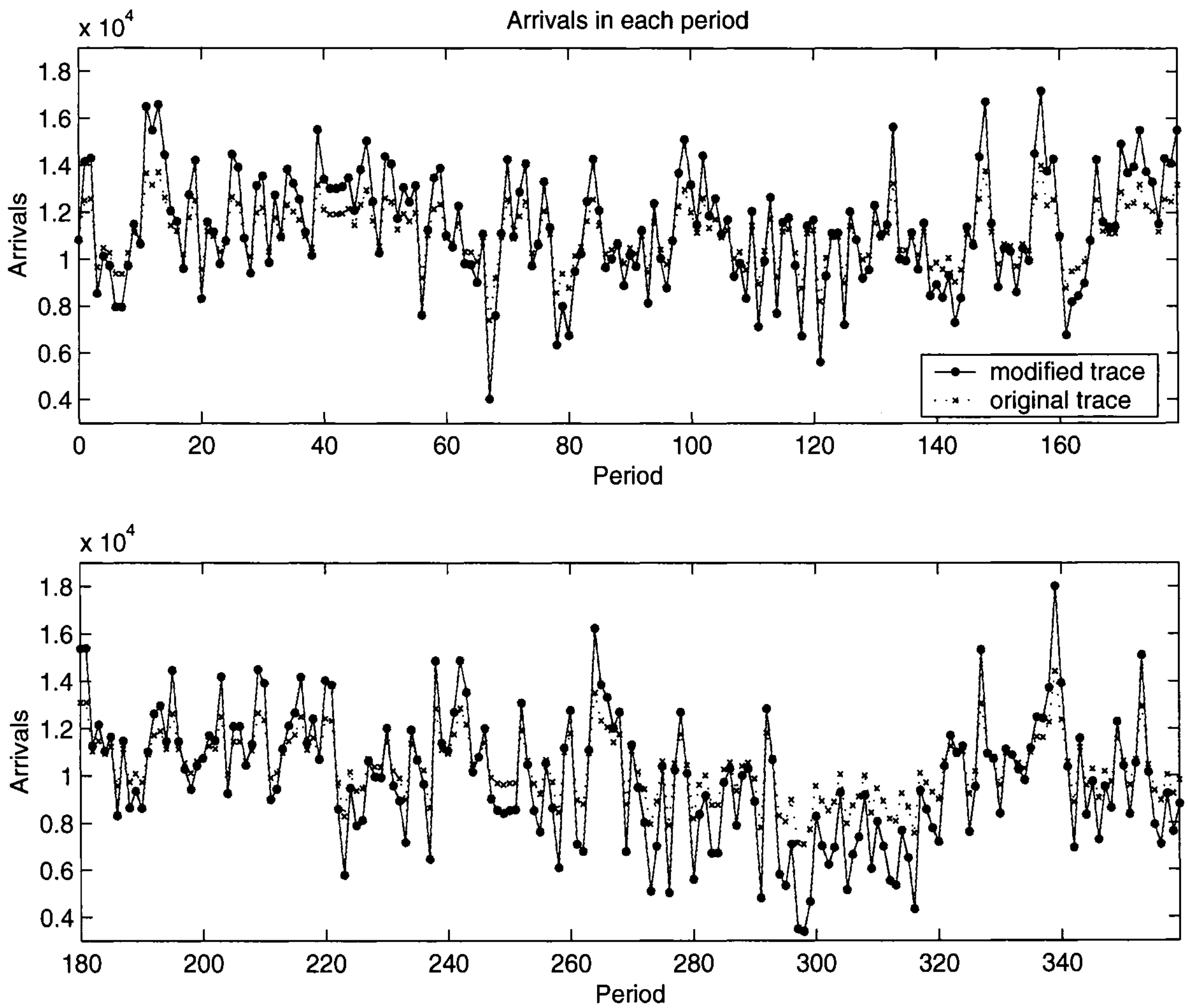

Figure 4.1: The arrivals in each period

4.2 Similarity Analysis

The modification on the traffic trace might change the arrival distribution pattern, leading to the change of the essence of the real trace, which is not desired. Thereby, similarity analysis is needed to verify whether the arrival distribution pattern of the modified traffic is similar to the one of the original traffic or not.

To compare the original and modified traces, the histogram of the arrival pattern of the two traces will be presented and analyzed. The time interval of any two consecutive packets is measured. The histogram of the original data trace is shown in figure 4.2. The histogram of the modified one is presented in figure 4.3. In these two figures, the x-axis represents the time interval of two consecutive packets and y-axis represents the amount of the packets having an interval indicated as x-axis. We can see that the arrival patterns of the original trace and the modified trace are similar, except that the time intervals are shrunk and the number of packets is increased. These are induced by the modifications we have done to the traffic.

It is then concluded that the similarity is reserved through the modifications. The modified trace can represent the original one and therefore is appropriate to be used in our project.

We also notice that these two traces are composed of irregularly distributed packets enveloped by an exponential distribution. Since we do no have the knowledge of how the packets were collected and what the packets were, there is no way to explain the phenomenon.

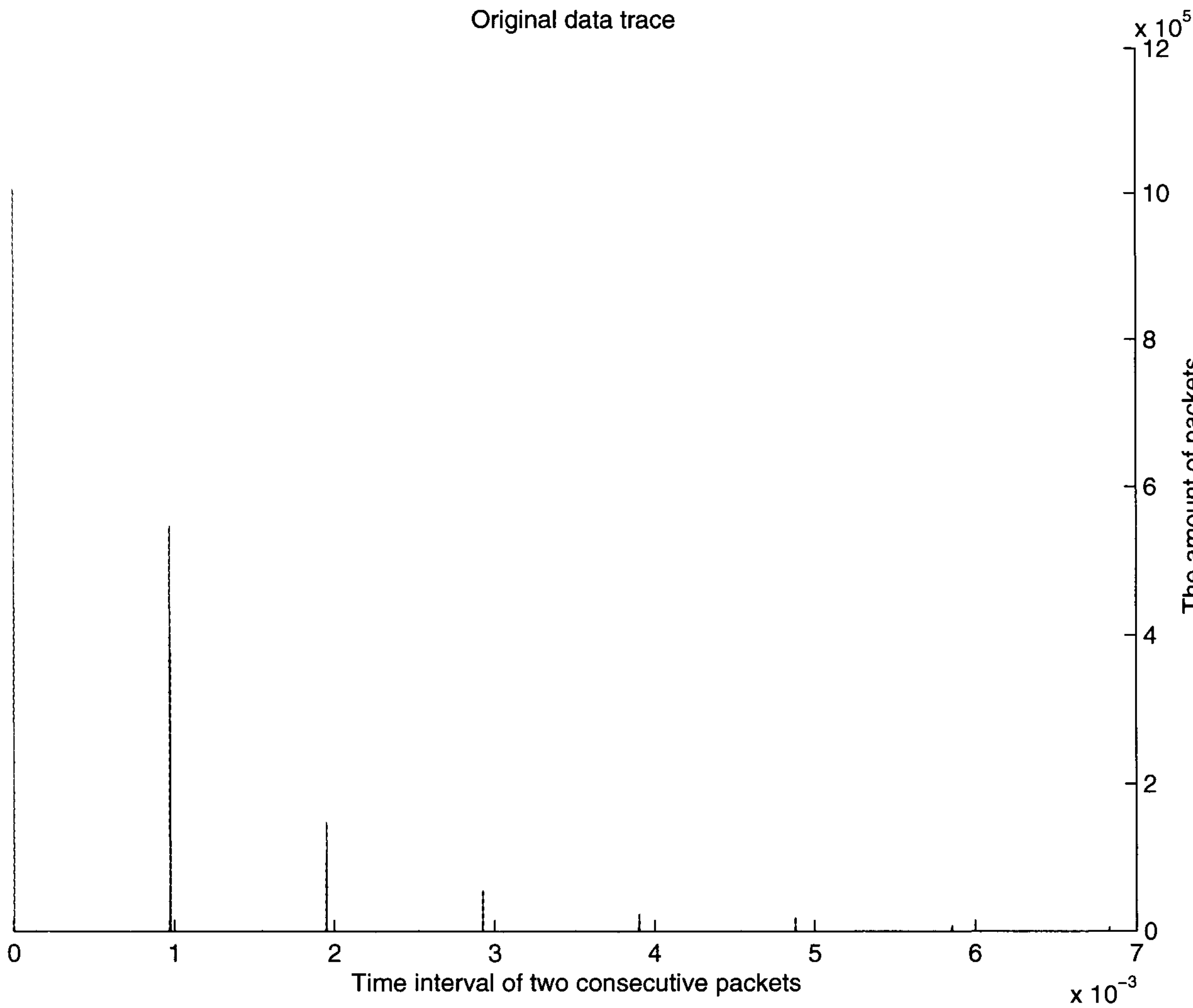

Figure 4.2: Arrival pattern of the original data trace

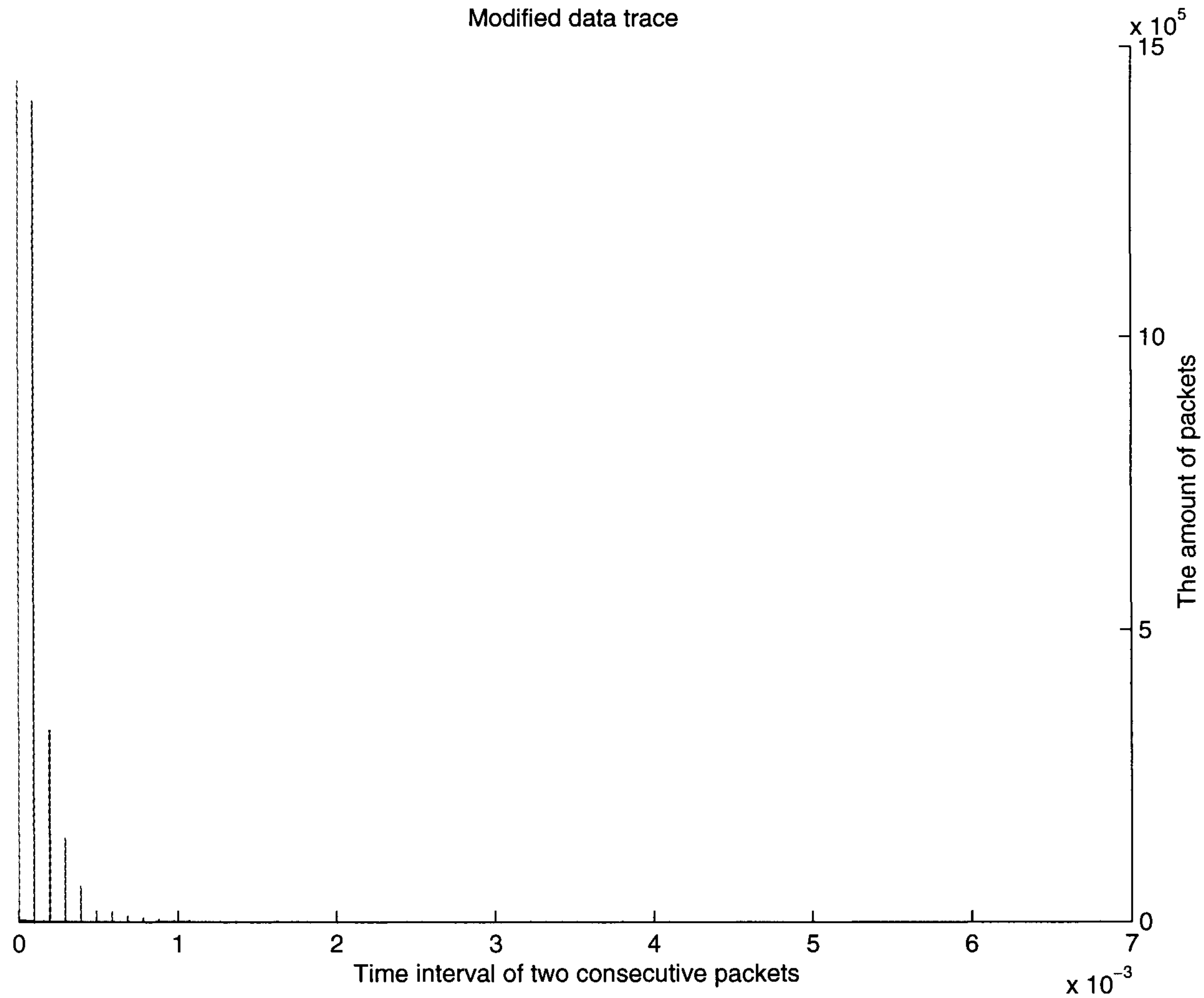

Figure 4.3: Arrival pattern of the modified data trace

Chapter 5

The Simulations

Simulations have been developed in order to test the allocation method proposed in this project. Only one connection will be considered in the simulation, as the allocation process is all the same to any connection. The real data trace discussed in chapter 4 will serve as the input traffic in the simulation.

5.1 The Simulation Model

In chapter 3, we mentioned that a period contains a number of TDM frames which in turn contains a set of time slots. The simulation model is shown in figure 5.1. A certain amount of time slots of every TDM frame is reserved for the traffic of a connection. The arriving packets of various lengths will be inserted into a queue if the buffer is not full. In the case that the buffer is full, the packet will be dropped. When a TDM frame arrives, the packets in the queue will be packed into the reserved time slots. Packets that have failed being packed will remain in the queue and will wait for another TDM frame.

The algorithm to implement the model is described in figure 5.2. When the simulation begins, we enter into the first TDM frame of the first period. The packet is read from the data trace and then it is inserted into a waiting queue if the buffer

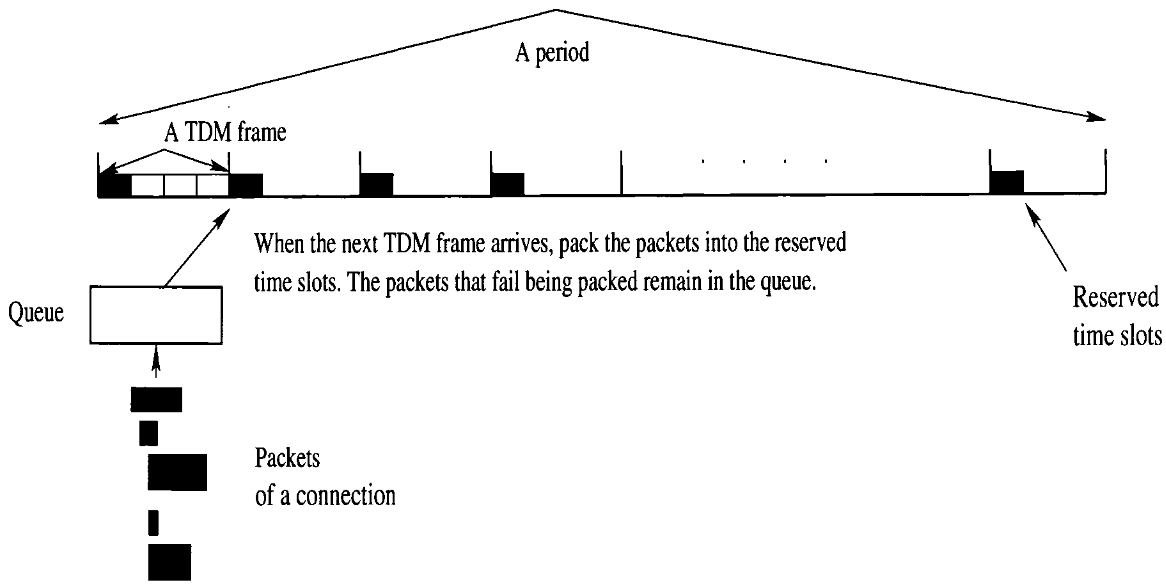

Figure 5.1: The simulation model

is not full. In the case that the buffer is full the packet will be dropped which contributes to the packet loss rate of this period. The reading and inserting process will be repeated until the current TDM frame ends.

When the current TDM frame ends, the packets in the buffer will be packed into the pre-reserved time slots and be transmitted. And then the next TDM frame of this period begins. This process repeats itself until the current period ends.

Once the current period ends, the packet loss rate in this period will be calculated and recorded, after which the assignment for the next period will be done according to the process that is described in section 3.4. Then we will pass to the next period. The whole process will be repeated until the data trace ends.

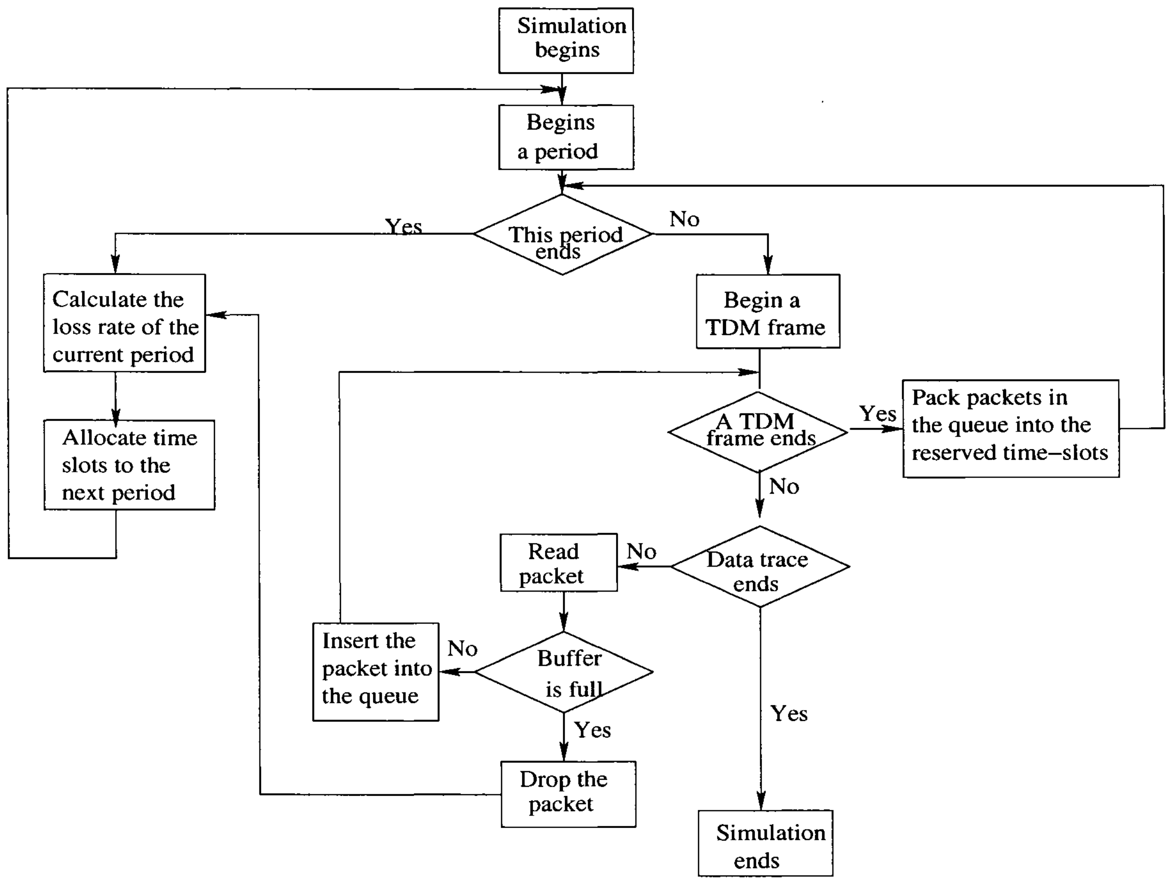

Figure 5.2: The simulation algorithm

5.2 The Simulation Platform

All the programs of the simulation for this project are coded in Java. The simulation is in part based on SSJ [1], which will be briefly introduced first. The classes especially developed are discussed in section 5.2.2.

5.2.1 SSJ

SSJ (Stochastic Simulation in Java) is a Java library for stochastic simulation, developed in the Département d'Informatique et de Recherche Opérationnelle (DIRO), at the Université de Montréal. It provides facilities for generating random variables, computing different measures related to probability distributions, and programming discrete-event simulations.

The class of *Resource*, which corresponds to a facility with limited capacity and a waiting queue, is used in our simulation. Several methods are modified to adapt to our requirements.

5.2.2 The Classes

Simulation_all_appro1 implements the main simulation of the first approach, while *Simulation_all_appro2* implements the second one. *Simulation_bollinger1* and *Simulation_bollinger2* simulate the allocation with BBs serving as the bounds. *Simulation_envelop1* and *Simulation_envelop2* simulate the allocation with MAEs. The main simulations and the simulations with BBs/MAEs will be discussed in chapter 6 and 7 respectively. Finally, *Simulation_individual* implements the fixed and the peak rate allocations.

PackQueueList aims to pack the packets in the buffer to the reserved time-slots. It is called in all the previous seven classes.

PackPackets, which is called in class *PackQueueList*, implements packing a single packet to the time-slots. This class takes advantage of the class *Resource* of SSJ platform. A time-slot is defined as an object of *Resource*.

DealTraffic implements the modifications to the real data trace mentioned in chapter 4. The data trace is a text file. The first column is the arrival time of a packet while the second column is the packet length. The input of the class is the original data trace *dec-pkt-4.tcp*, the output is the modified data trace.

GetArrivals collects some features of a data trace, such as the total packets, packets in each period and total bytes in the data trace.

GetPeakRate will get the peak arrival rate (bytes/TDM frame) in a TDM frame.

5.3 Parameters

The parameters used in the simulation are summarized in this section.

The rate of a wavelength F is 1.6 Gbps;

The duration of a time slot is 1 μs;

$\Rightarrow$ the length of a time slot $L_s = (\frac{1.610^9}{8} \times 1 \times 10^{-6}) = 200$ bytes

The duration of a period is 1 s;

The duration of a TDM frame is 200 μ s

$\Rightarrow$ a TDM frame contains 200 time slots ($S_f = 200$).

$\Rightarrow$ a period contains 5000 TDM frames ($M = 5000$).

The mean packet length $L_p = 285$ bytes;

The mean number of arrivals in a period $\bar{A} = 10637$;

There are 360 periods in the trace, therefore $k = 360$.

The average packet loss rate (QoS requirement) is $\leq 2\%$. The acceptable range is then defined as [1.2%, 2.3%], and R_t^{++} and R_t^{--} are 4% and 0.8% respectively. Thereby, c are defined as $c_0 = 20$, $c_1 = 100$ and $c_2 = 40$.

The parameters that we have chosen in the performance evaluation equation (3.17) are: $P_0 = 0.4,\ P_1 = 0.3,\ P_2 = 0.2,\ P_3 = 0.1$. The values of P are determined according to the importance of the items. For instance, P_0 has the highest value because the time slots utilization efficiency ε is mostly considered in this project. γ defined in penalty function 3.16 is set to 200.

7 time slots are assigned to the first period. The initial allocation could be the adjustment result of the former sojourn time, an average value of the history or even a random value in a reasonable range.

As mentioned in 1.2, the traffic is assumed to be of a certain type for a real time application. Buffers containing at most 5 packets are defined in the simulation so as to decrease the delay.

5.4 Cases with various weights

As described in section 3.5, the calculation of the weighted assignment and the estimation of the packet loss rate are done based on the historical data. To be more precise, these values are obtained in a weighted way.

6 cases with different weights are considered in the simulation. All the 6 cases will be explored under the two approaches mentioned in section 3.5. They are listed in table 5.1.

The weights are identically allocating in both approaches. An increasing number of previous periods is taken into account in slots, from only 1 to 3 and then to 5. For cases 2, 3 and 4 of each approach, in which three previous periods are taken into account, the weights are distributed in an increasing uniform way. The same thing happens to the cases 5 and 6 in which five previous periods are taken into account. We want to find out not only the appropriate number of previous periods that we should trace back but also the distribution of weights that will give the best system performance.

Table 5.1: Different cases under consideration

Approach		Case	Weights associated				
			α_1	α_2	α_3	α_4	α_5
1:	$W_{i+1} = N_i$ $E_{i+1} = \alpha_1 R_i + \alpha_2 R_{i-1} + ...\alpha_n R_{i-n+1}$ $\sum_j \alpha_j = 1 \qquad j = 1, ...n$	1-1	1	0	0	0	0
		1-2	0.8	0.1	0.1	0	0
		1-3	0.6	0.3	0.1	0	0
		1-4	0.4	0.3	0.3	0	0
		1-5	0.6	0.1	0.1	0.1	0.1
		1-6	0.3	0.3	0.2	0.1	0.1
2:	$W_{i+1} = \alpha_1 N_i + \alpha_2 N_{i-1} + ...\alpha_n N_{i-n+1}$ $E_{i+1} = \alpha_1 R_i + \alpha_2 R_{i-1} + ...\alpha_n R_{i-n+1}$ $\sum_j \alpha_j = 1 \qquad j = 1, ...n$	2-1	same as 1-1, ignored				
		2-2	0.8	0.1	0.1	0	0
		2-3	0.6	0.3	0.1	0	0
		2-4	0.4	0.3	0.3	0	0
		2-5	0.6	0.1	0.1	0.1	0.1
		2-6	0.3	0.3	0.2	0.1	0.1

5.5 The Confidence Interval

A confidence interval is an interval used to estimate the likely size of a population parameter. It gives an estimated range of values that has a specified probability of containing the parameter being estimated. In other words, a confidence interval can be represented as a range Δ, such that $Pr\{\bar{M} - \Delta < x < \bar{M} + \Delta\} = 1 - a$, where $\bar{M}$ = the estimate mean, Δ = standard error, $1 - a$ = the level of the confidence interval.

A confidence interval is based on three elements:

1. the value of a statistic
2. the standard error of the measure
3. the desired width of the confidence interval, e.g. 95% or 99%.

In our project, the confidence interval of the mean of the packet loss rate is considered, and the corresponding value is 95%. The formula is as follows:

$$\Delta = \left[M - 1.96\frac{\delta}{\sqrt{n_s}},\ M + 1.96\frac{\delta}{\sqrt{n_s}} \right] \tag{5.1}$$

where δ means the the standard deviation, and n_s is the sample number. In the simulation, the confidence interval of the mean packet loss rate will be calculated.

To conclude, in this chapter, the simulation model and the platform have been described. The parameters and the cases with different n and α were listed as well. The results of the main simulations and the simulations with bounds will be presented and discussed in chapter 6 and 7.

Chapter 6

The Main Simulations

6.1 The Results of the Main Simulation

Two approaches shown in table 5.1 were explored in the main simulations. Figures 6.1 to 6.6 show the time slots assignment result and the packet loss rate of the 6 cases of the 1^{st} approach mentioned in the former section, while 6.7 to 6.11 show the results of the 5 cases of the 2^{nd} approach.

We can see that the performance around the 300^{th} period is poor in the majority of the cases. In the case 1-5, 1-6 and 2-2 (see table 5.1) the packet loss rate even soar up to 99%. The reason for this phenomenon is further discussed in section 6.3.

Table 6.1 shows the comparison of the time slots utilization efficiency, the bandwidth occupation ratio, the average loss rate, the standard deviation from the QoS threshold as well as the performance evaluation value among all the cases. From that table, we can draw the following conclusions:

1. The average packet loss rate in all the cases satisfies the required QoS, i.e. 2%.

2. Generally speaking, the cases of the 2^{nd} approach work better than those of the 1^{st}. That is to say, it is better to weight not only the packet loss rate but also the time slots assignment of the previous periods.

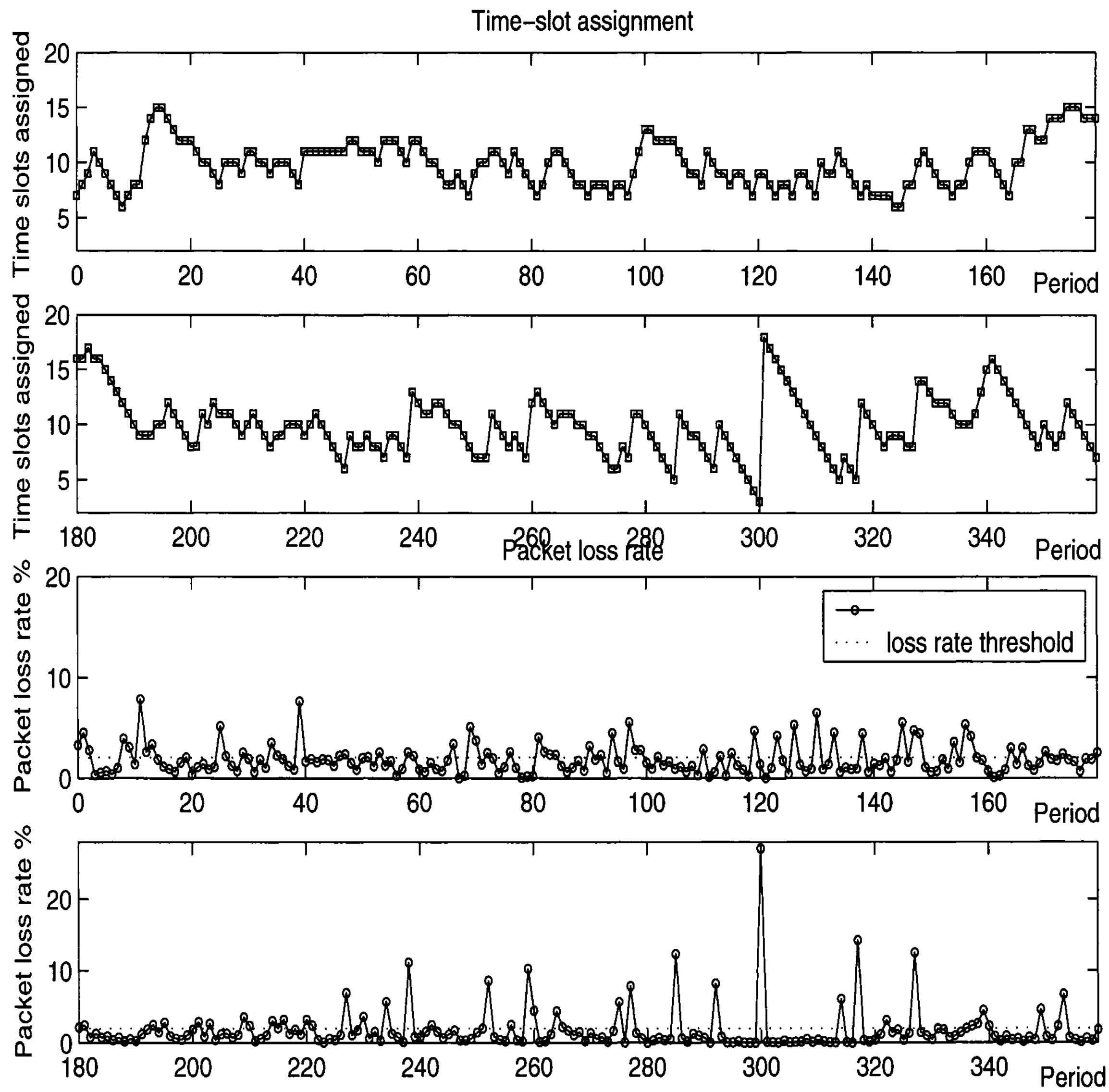

Figure 6.1: Time slots allocation and packet loss rate for case 1-1

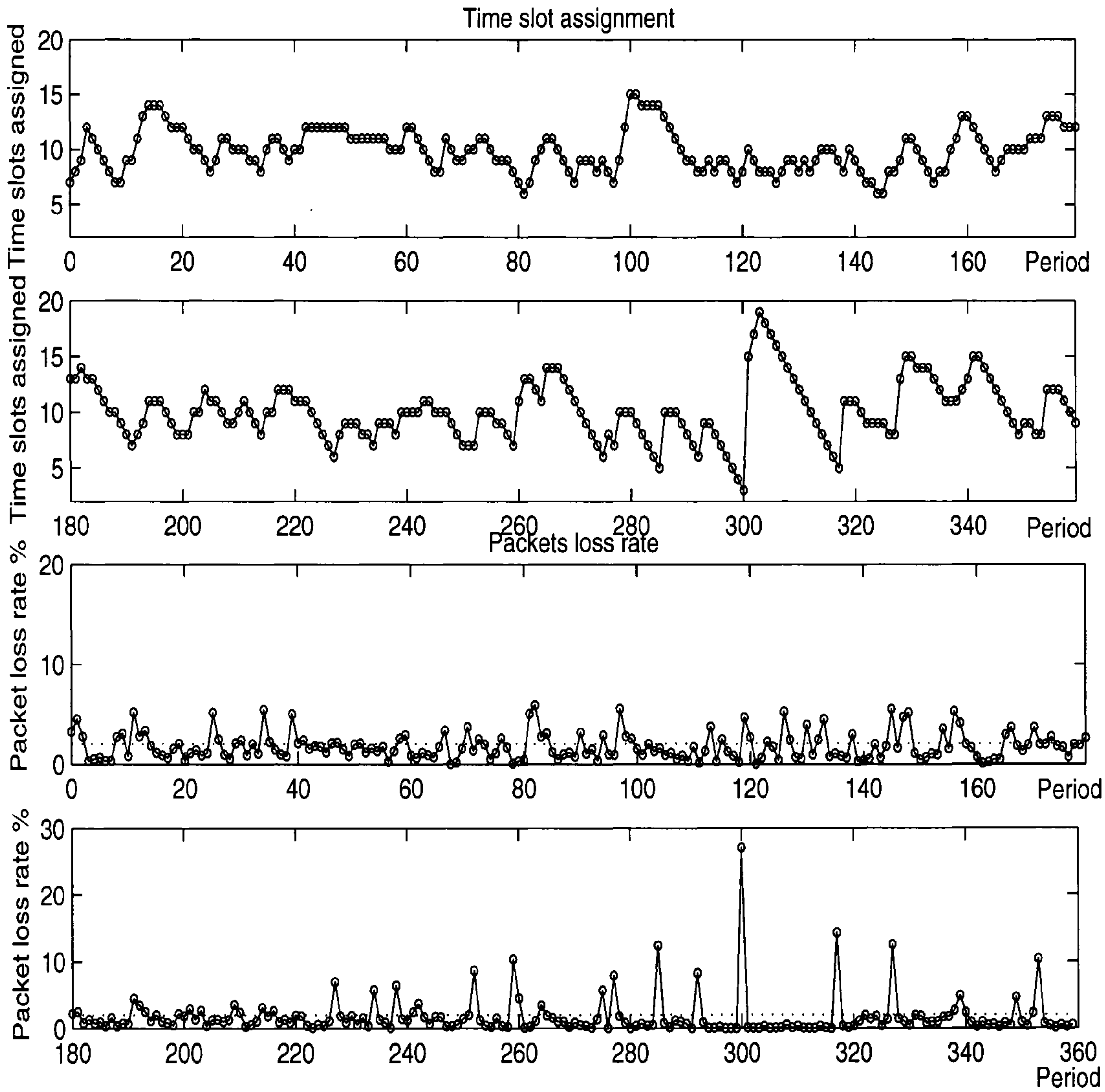

Figure 6.2: Time slots allocation and packet loss rate for case 1-2

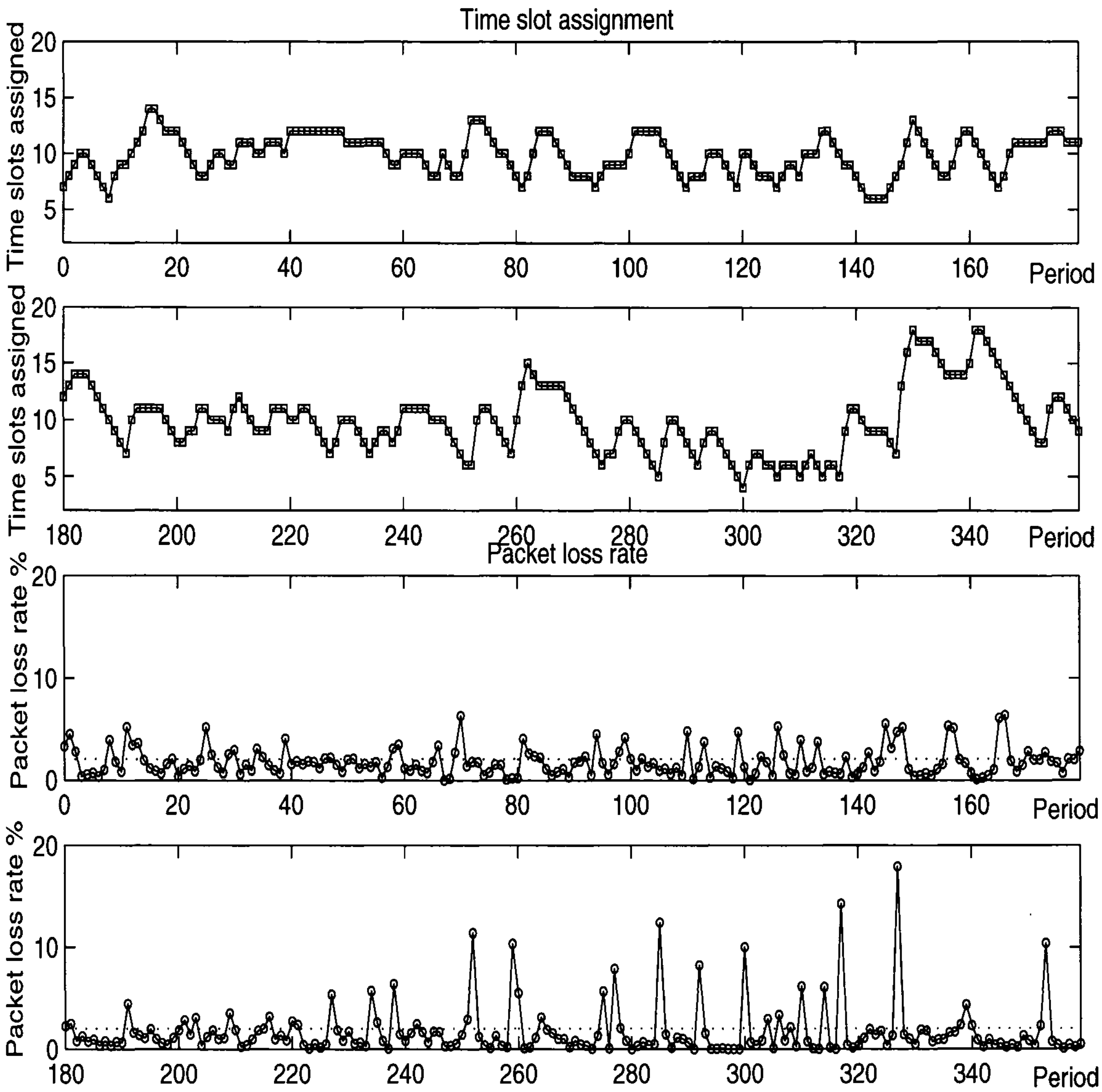

Figure 6.3: Time slots allocation and packet loss rate for case 1-3

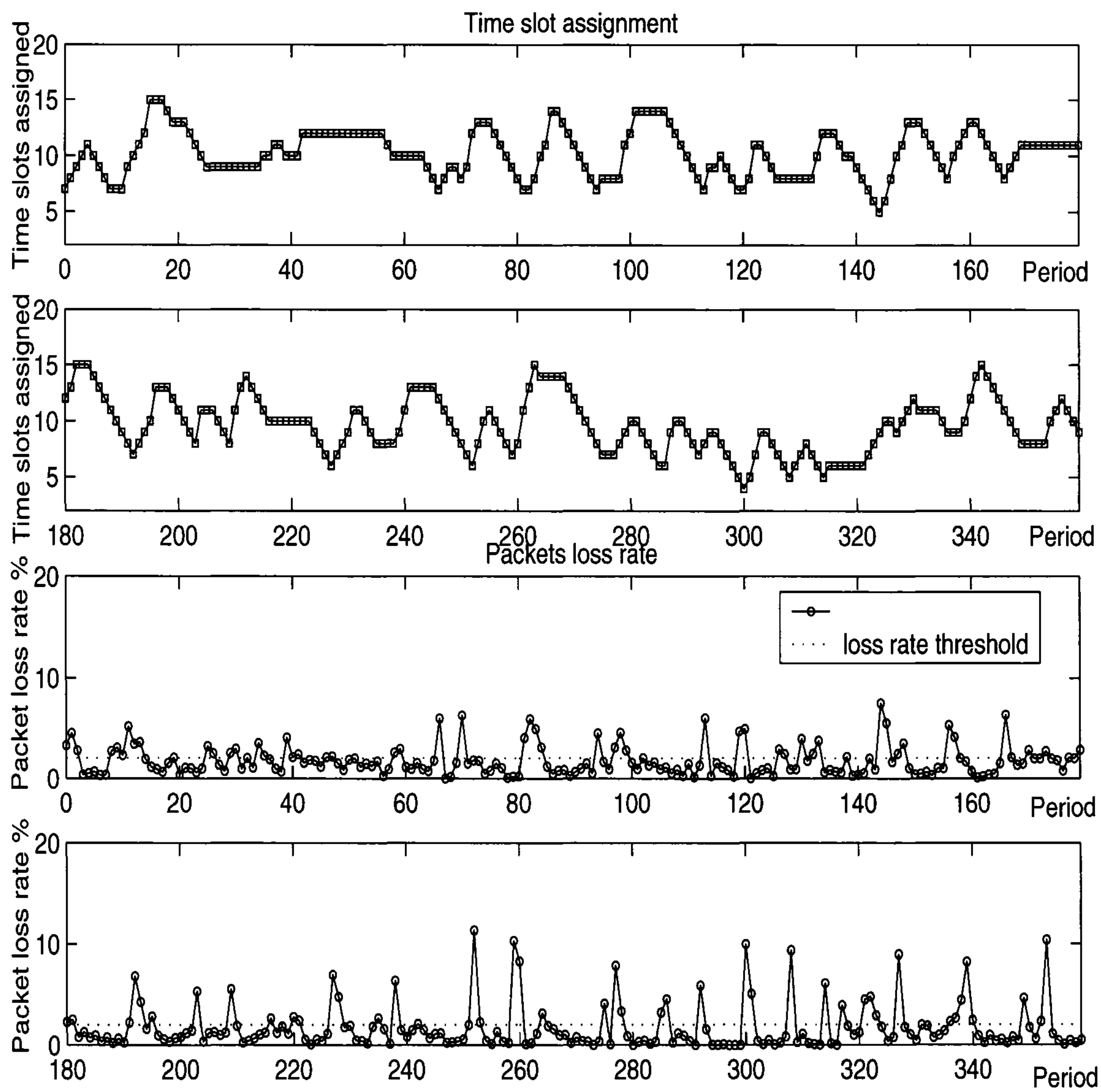

Figure 6.4: Time slots allocation and packet loss rate for case 1-4

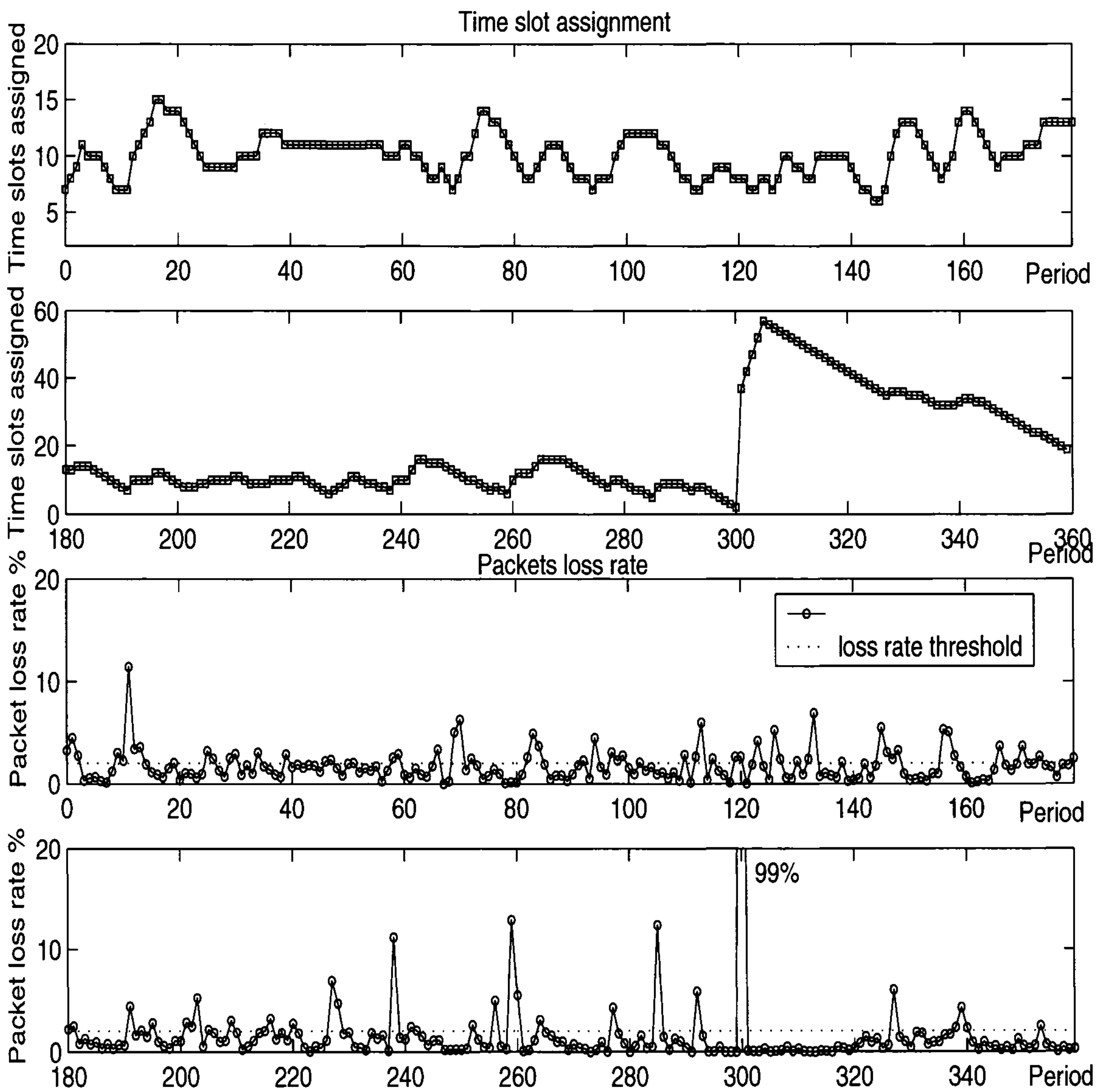

Figure 6.5: Time slots allocation and packet loss rate for case 1-5

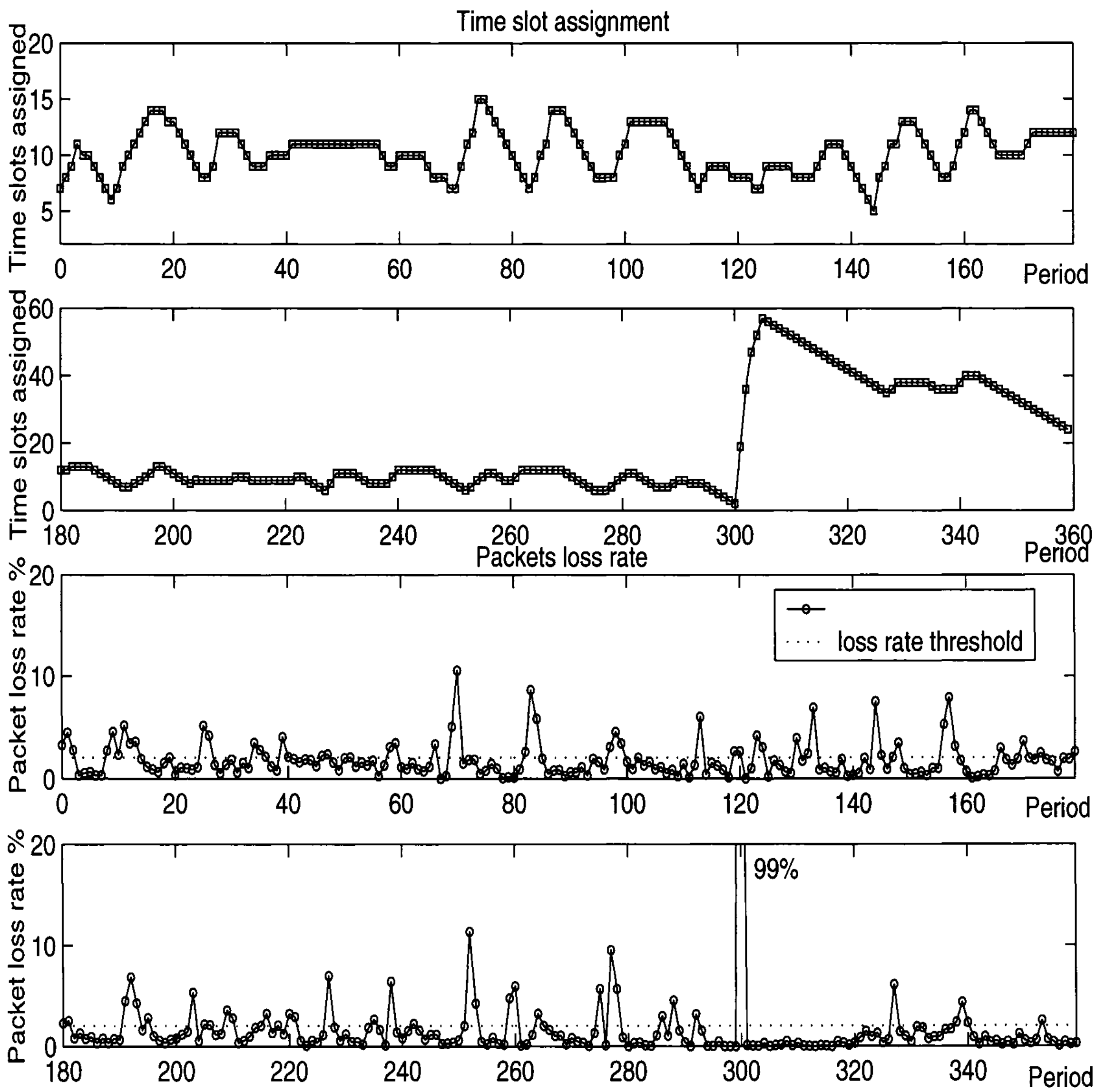

Figure 6.6: Time slots allocation and packet loss rate for case 1-6

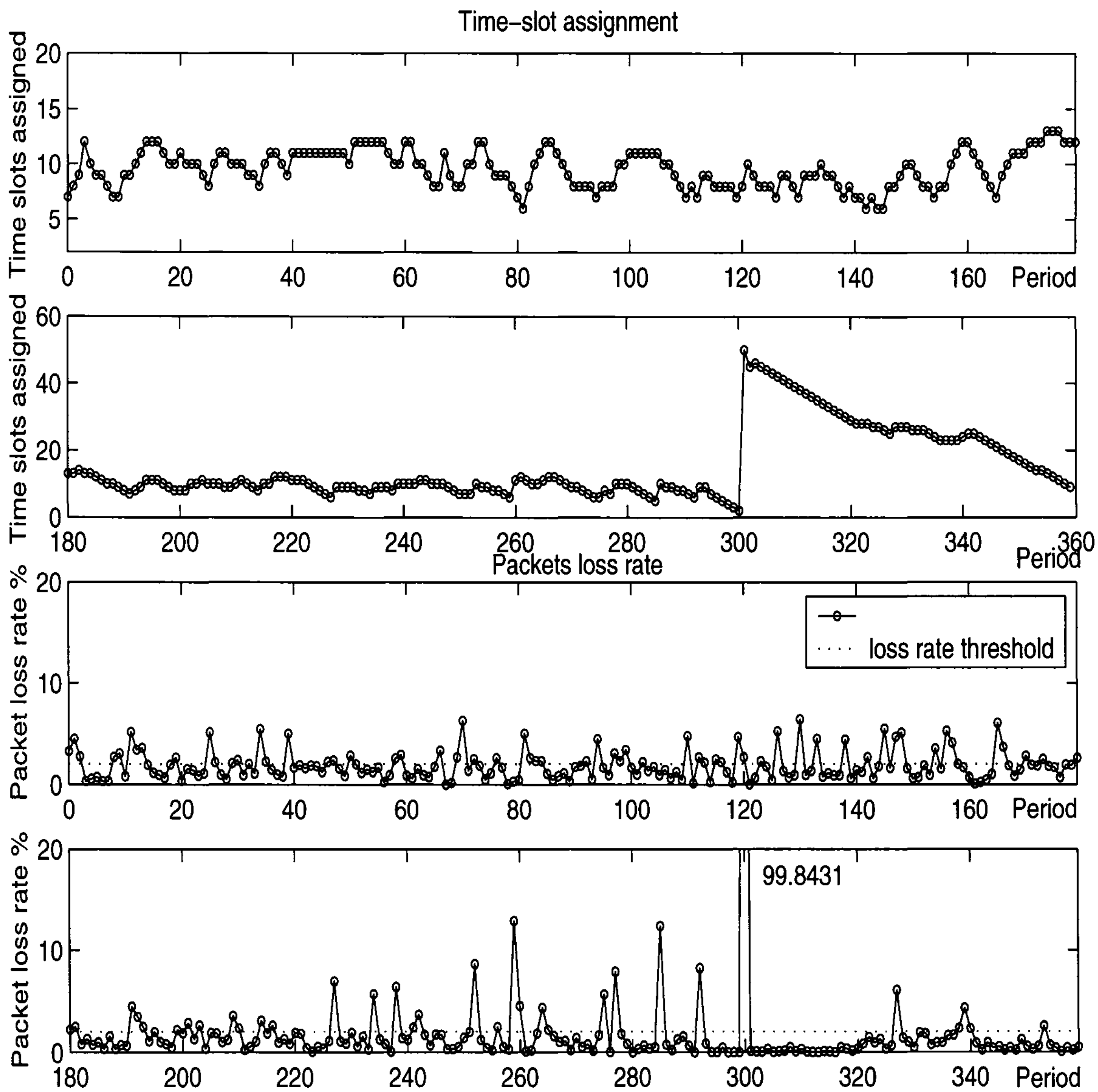

Figure 6.7: Time slots allocation and packet loss rate for case 2-2

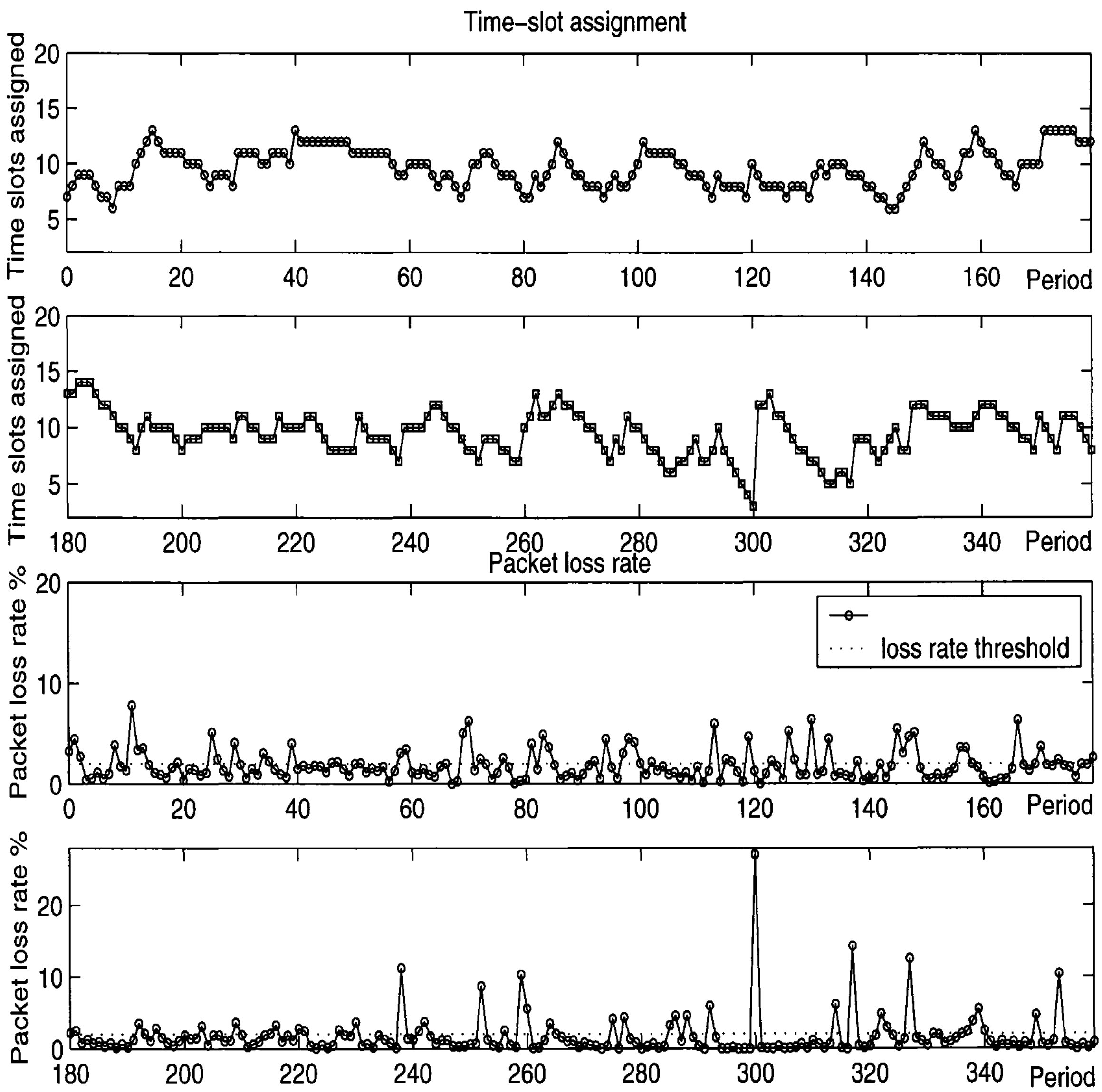

Figure 6.8: Time slots allocation and packet loss rate for case 2-3

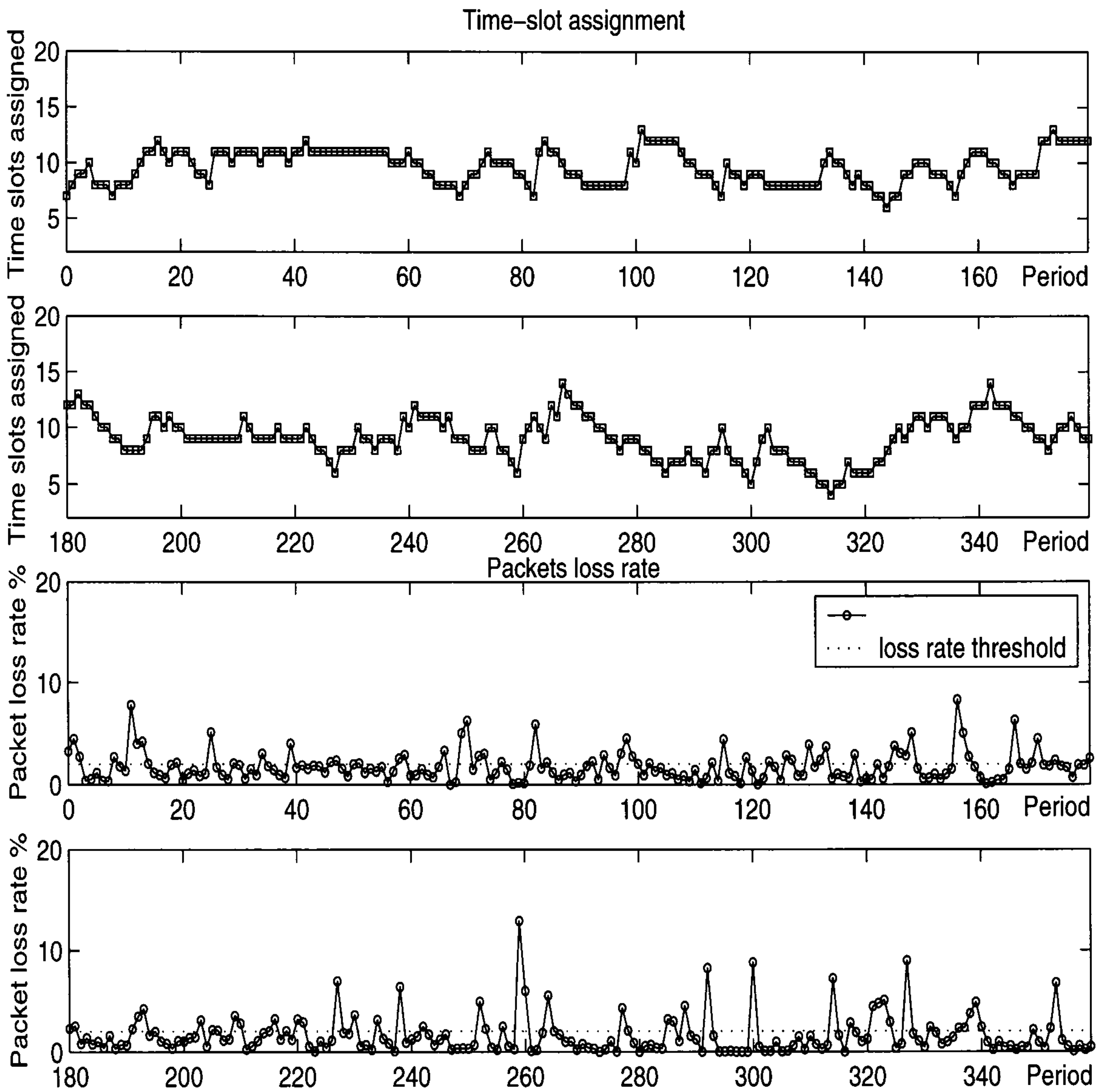

Figure 6.9: Time slots allocation and packet loss rate for case 2-4

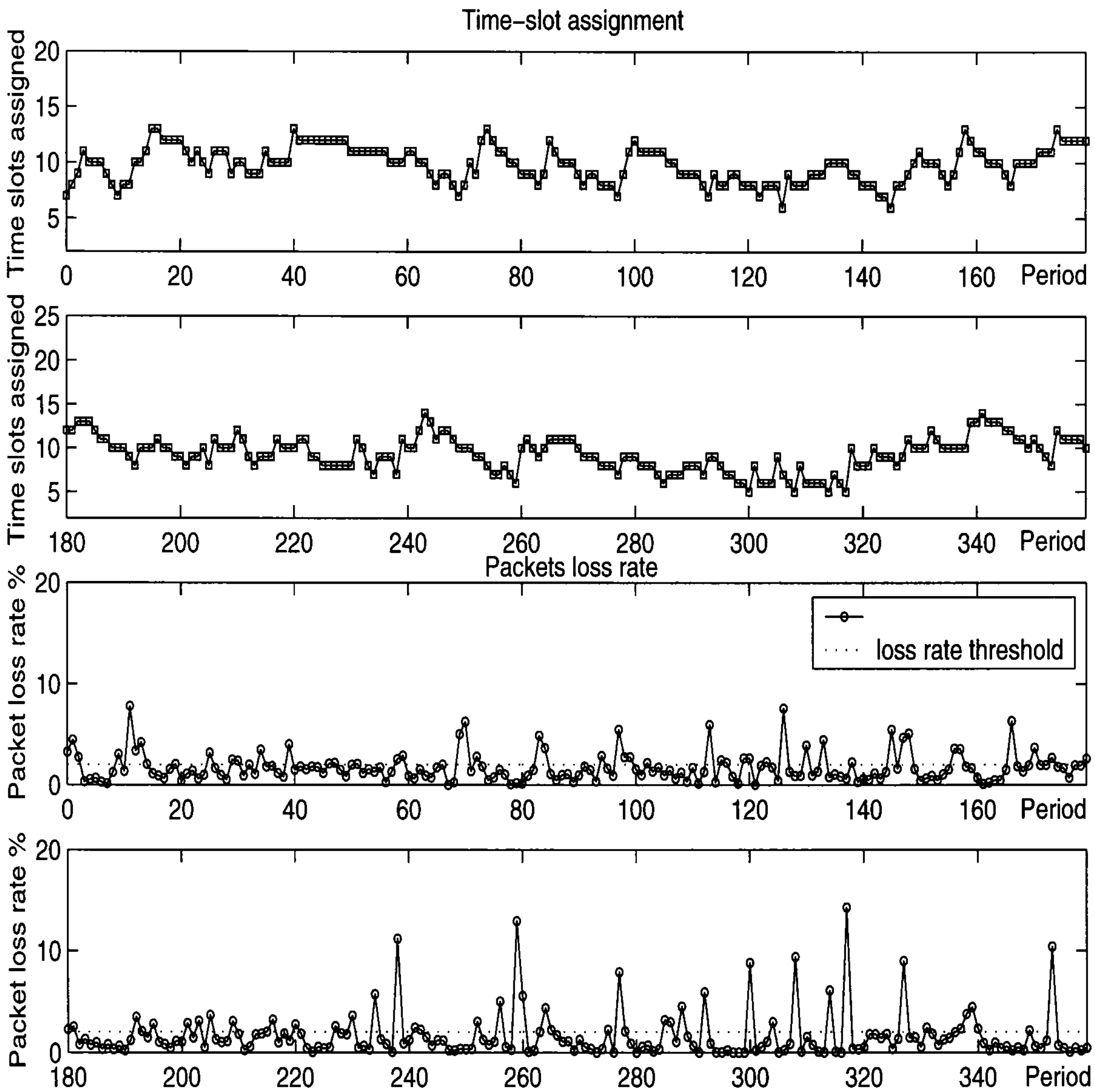

Figure 6.10: Time slots allocation and packet loss rate for case 2-5

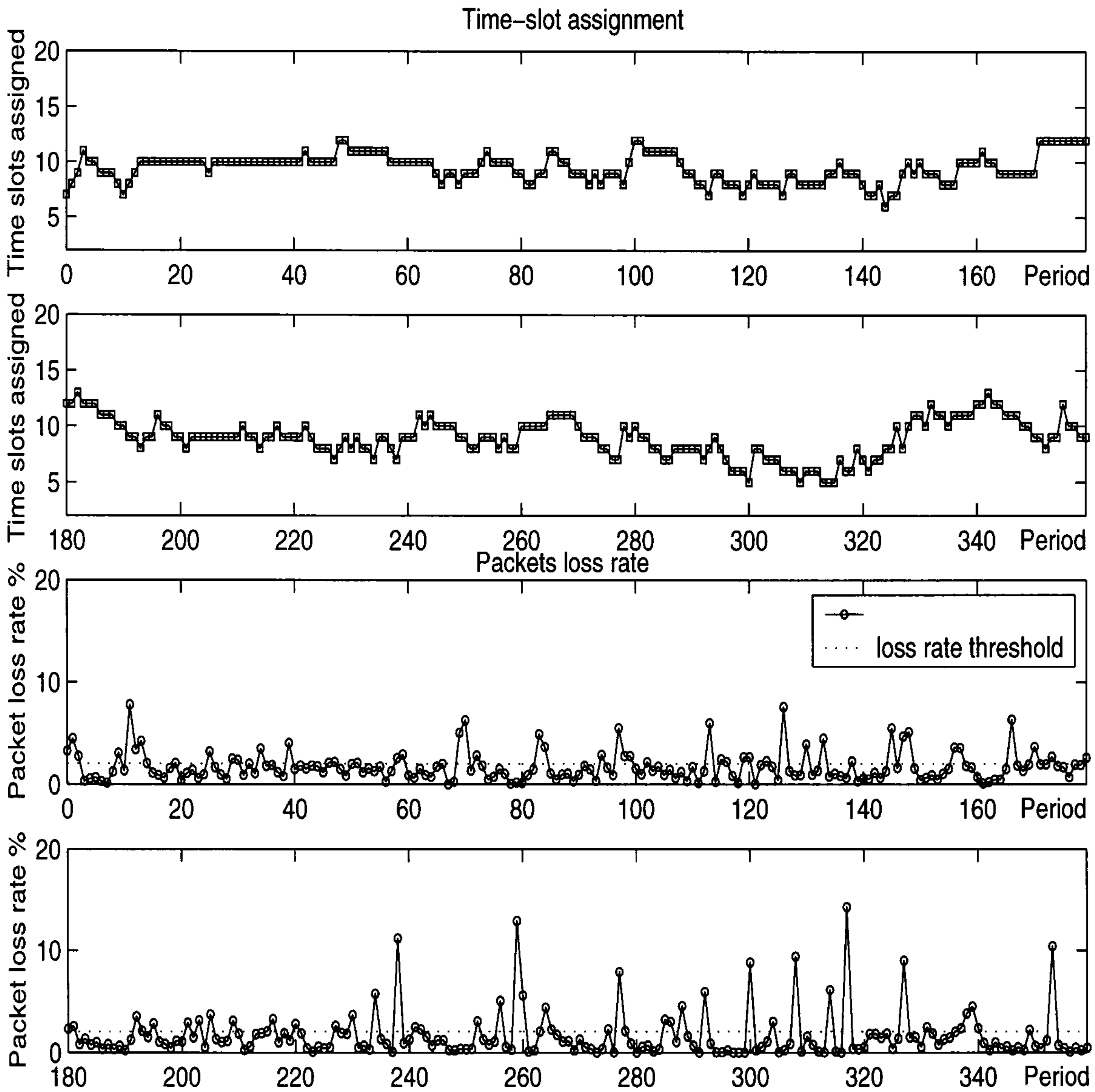

Figure 6.11: Time slots allocation and packet loss rate for case 2-6

Table 6.1: Performance comparison – main simulation

	Case	ε %	$\bar{R}$%	η%	V%	Z
1^{st} approach	case 1-1	15.23±0.70	1.81±0.2	4.94±0.12	2.38	4.13
	case 1-2	15.02±0.72	1.76±0.2	5.01±0.12	2.14	4.03
	case 1-3	15.21±0.70	1.77±0.2	4.95±0.12	214	4.17
	case 1-4	15.14±0.71	1.77±0.2	4.95±0.12	1.95	4.17
	case 1-5	10.24±0.80	1.83±0.5	7.15±0.55	5.46	0.89
	case 1-6	10.39±0.81	1.86±0.5	7.36±0.58	5.45	0.80
2^{nd} approach	case 2-2	12.23±0.76	1.93±0.5	6.16±0.40	5.44	1.96
	case 2-3	15.74±0.70	1.79±0.2	4.78±0.09	2.32	4.40
	case 2-4	16.09±0.70	1.72±0.2	4.68±0.09	1.73	4.69
	case 2-5	15.75±0.69	1.71±0.2	4.78±0.09	1.94	4.48
	case 2-6	16.33±0.71	1.73±0.2	4.61±0.08	1.69	4.81

3. Case 2-6 and 2-4 of the 2^{nd} approach have the best performance evaluation values. The average loss rates are 1.73% and 1.72% respectively.

4. When several periods are taken into account, it is better to distribute the weights in an uniform way to those periods like the case 2-4 and 2-6, instead of putting too much emphasis on the recent one.

Figure 6.12 shows the histogram of time slots assigned to each period. The X-axis is the amount of the time slots assigned and the Y-axis is the percentage. The histogram represents the distribution of the assignment, and helps reveal the bandwidth occupation status to some extent. For instance, a high percentage of many slots assigned might mean a bandwidth waste, and a high percentage of few slots assigned might imply a high packet loss rate.

Comparing the histogram, we can observe that:

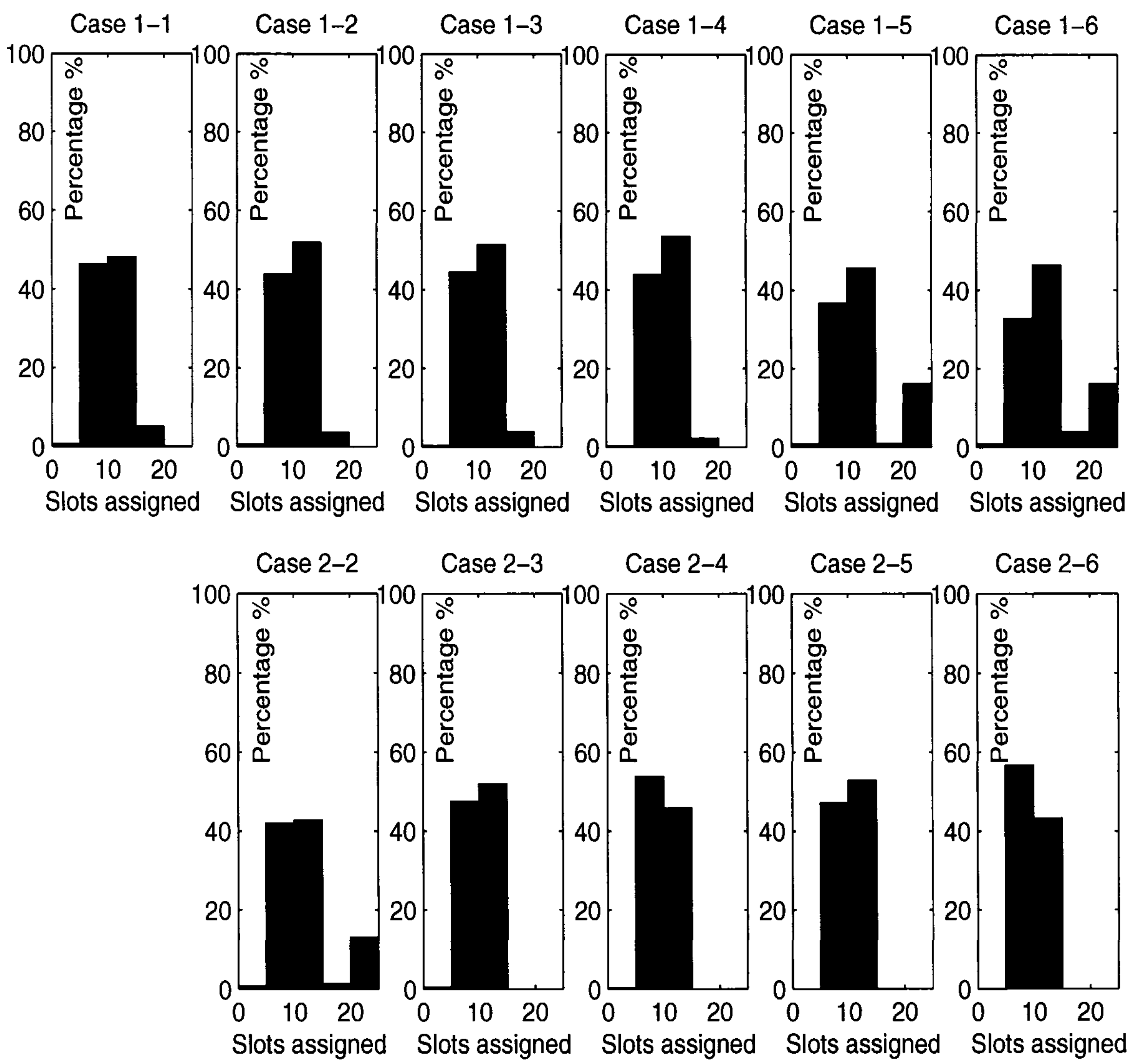

Figure 6.12: The histogram of the time slots assigned – main simulation

1. Except the case 2-2, all the other cases of the 2^{nd} approach have no allocation larger than 15 time slots. This means except the case 2-2, the 2^{nd} approach has a low bandwidth occupation ratio, which justifies the conclusion that most cases in 2^{nd} approach work better than those of the 1^{st}.

2. We can see that case 2-4 and 2-6 have less allocations between 10 and 15. This indicates a low bandwidth occupation ratio of case 2-4 and 2-6, which complies with the conclusion that case 2-4 and 2-6 work best and thereby have the best performance evaluation values.

6.2 Comparison with Conventional Reservation Methods

In this section, the simulation results are compared with two conventional reservation methods: fixed allocation and reservation based on the peak arrival rate.

In fixed allocation, just as the name implies, the number of time slots assigned to each period is fixed to a certain value. We mentioned that in the simulation with dynamic allocation, 7 time slots are assigned to the first period. In order to compare the performance of the two allocation mechanisms, the fixed allocation is done with 7 time slots.

It is worth mentioning that the biggest disadvantage of the fixed allocation is that there is no adjustment ability. If the value assigned is not an appropriate one, the performance will be very poor. The packet loss rate of the fixed allocation is shown in figure 6.13.

Another allocation strategy is to reserve the time slots according to the peak arrival rate of a connection. The peak arrival rate is usually obtained from statistics.

Through a small program, we measured the peak arrival rate which was 4306 bytes/TDM frame. Since a time slot can only contain 200 bytes, 4306 bytes will

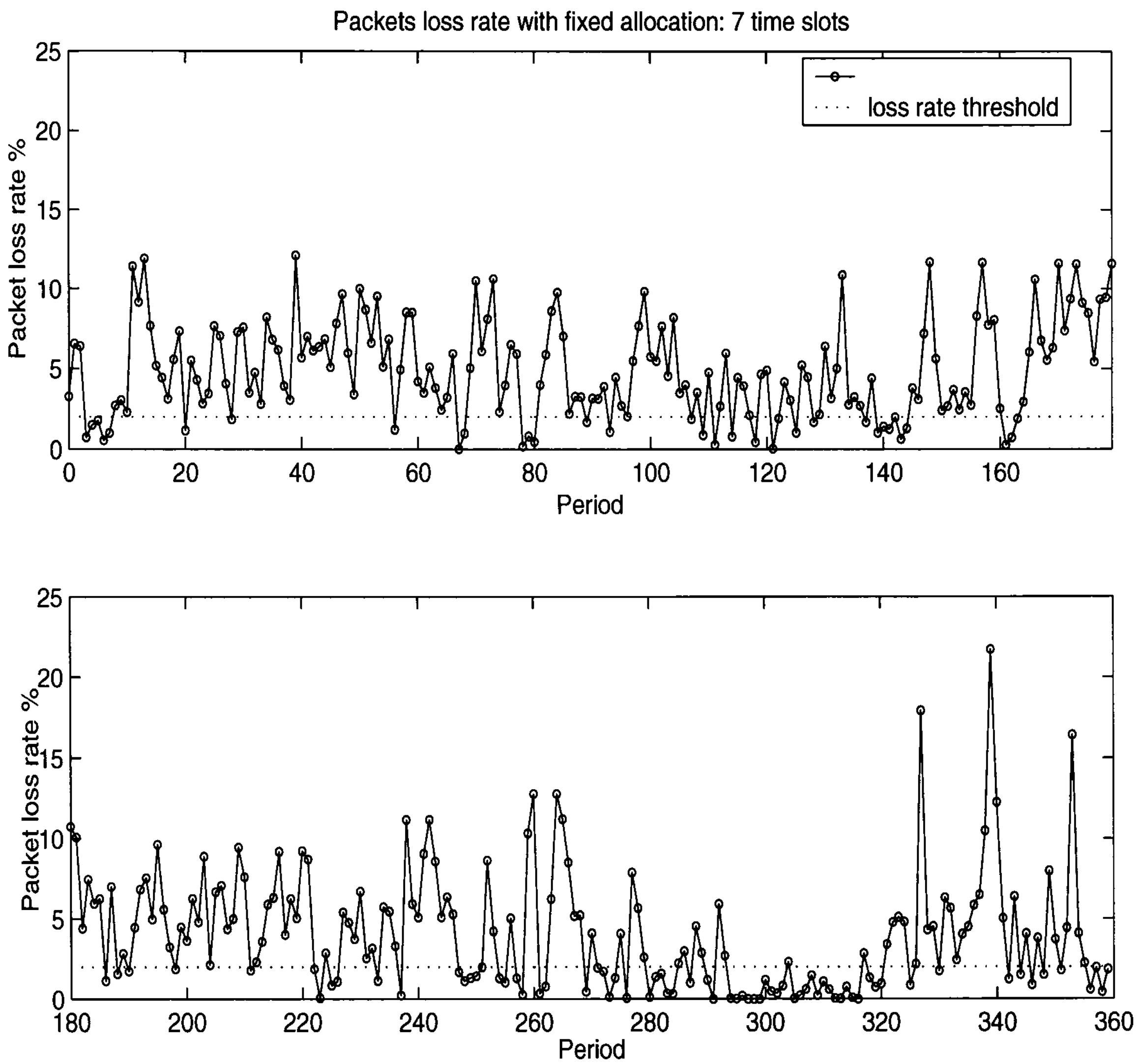

Figure 6.13: Packet loss rate with fixed allocation (7 time slots)

then need $4306/200 = 21.53 \approx 22$ time slots. In this simulation, 22 time slots are assigned to the connection.

The packet loss rate of the peak rate allocation is showed in figure 6.14.

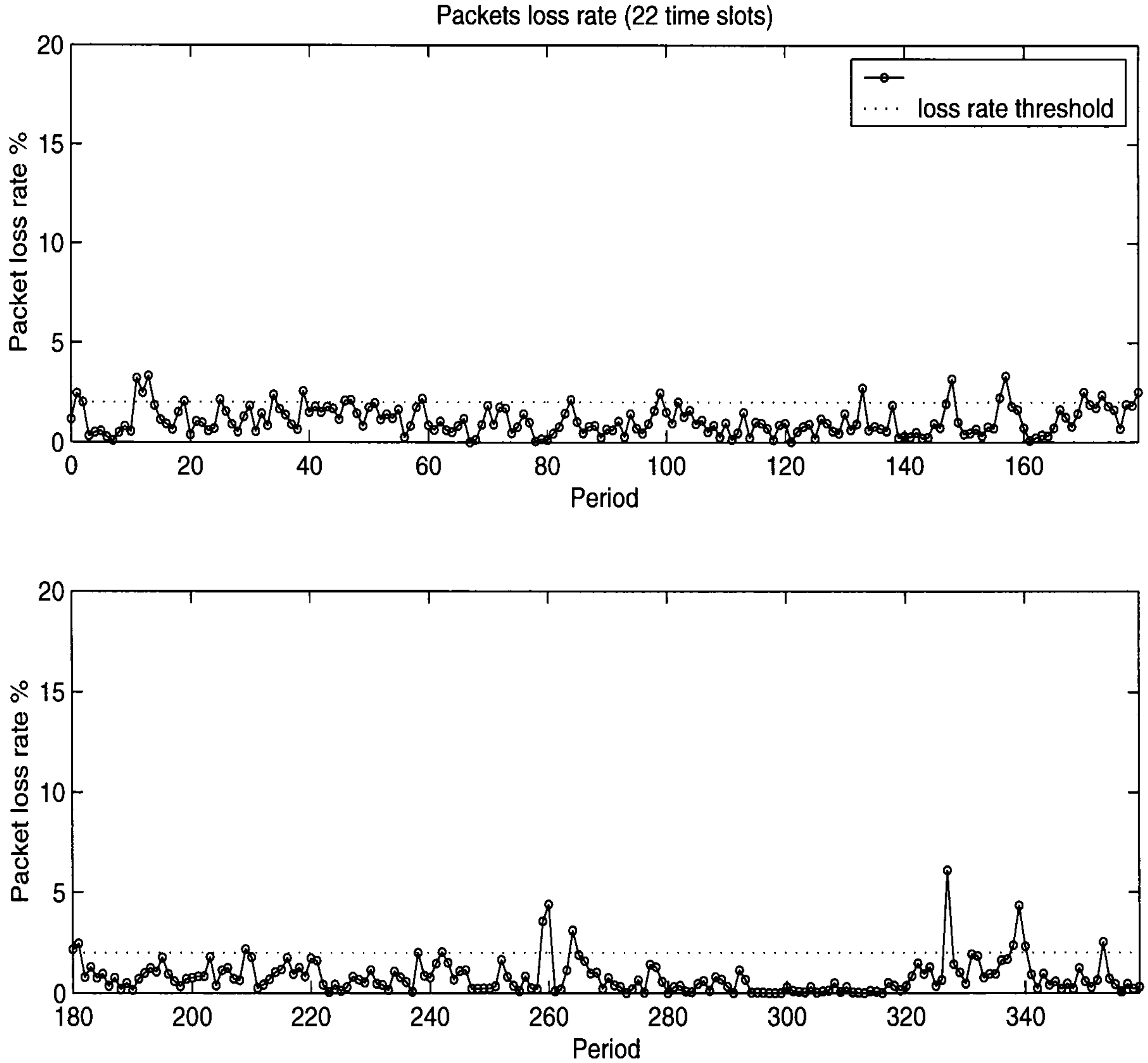

Figure 6.14: Packet loss rate with fixed allocation (22 time slots)

Table 6.2 compare the performance between the two conventional methods with the dynamic one. We can see that the fixed allocation has a very high packet loss

Table 6.2: Performance comparison with conventional allocation methods

Case	ε %	$\bar{R}$%	η%	V%	Z
case 2-4	16.09±0.70	1.72±0.2	4.68±0.09	1.73	4.69
case 2-6	16.33±0.71	1.73±0.2	4.61±0.08	1.69	4.81
Fixed allocation (7 time slots)	20.07±0.86	4.48±0.4	3.50±0	4.21	-13.86
Peak rate Allocation (22 time slots)	6.96±0.32	0.95±0.08	11±0	1.33	-0.78

rate $\bar{R}$, while the peak rate allocation has a low time slots utilization efficiency ε and a high bandwidth occupation ratio η. Therefore, their performance evaluation values are very low. Case 2-6 of the dynamic allocation has the best performance.

6.3 Transient Analysis

From the figures above, we notice that most cases have a problem in period 300, where the packet loss rate is extremely high. Consequently, the time slots assigned to the following periods will be too large. The reason leading to the phenomenon will be discussed in this section.

Consider the simulation of case 2-2 as an example. Table 6.3 and figure 6.15 show what happened around period 300. First of all, it is worth mentioning that the arrivals in periods before the 300^{th} period are comparatively less than that in the other periods. Also notice that the arrivals increase in period 300.

From table 6.3, we can see that in period 296 the estimated loss rate of period 297 is 0.36%, inferior than the lower bound of the acceptable range [1.2%, 2.3%]. Therefore, fewer time slots are assigned to period 297. This process repeats up to period 299. 5 slots are assigned to period 297, 4 slots are assigned to period 298, 3 slots are assigned to period 299. In period 299, the estimated loss rate of period 300

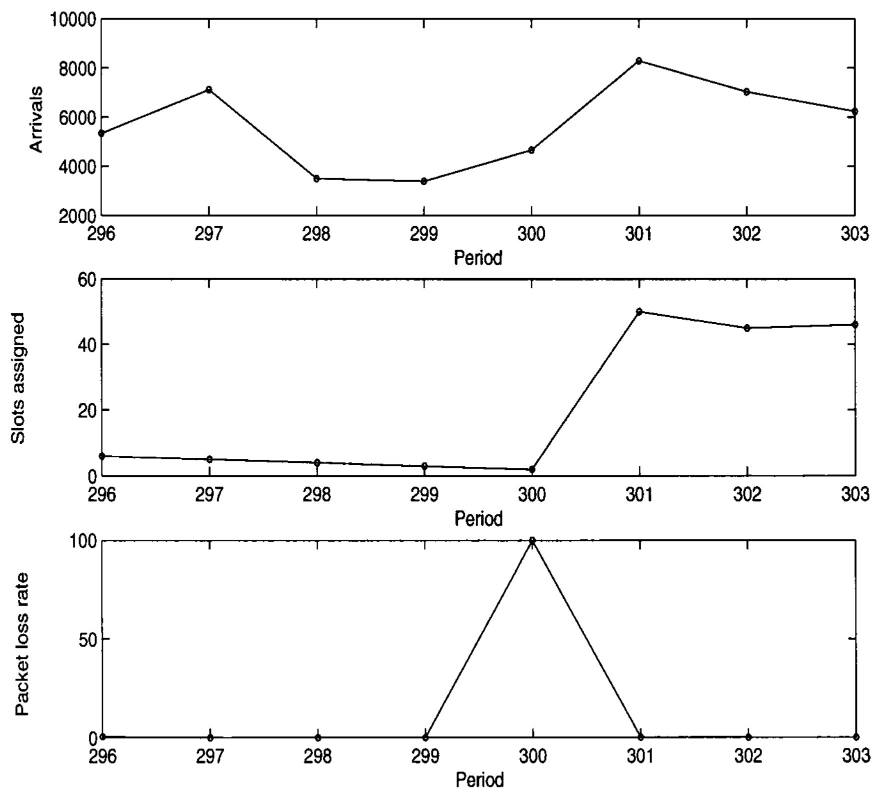

Figure 6.15: What happened around period 300

is even lower, i.e. 0%. Consequently, one more time slot is eliminated, i.e. only 2 time slots are assigned to period 300. Unfortunately, the traffic in period 300 actually increases. With less time slots and more traffic, it is not surprising that the packet loss rate of period 300 soars up to 99.84%. It is quite natural that the estimated loss rate of the following periods will be high as well, since we have a high loss rate in period 300. This will lead to a jump of time slots assigned to the following periods, as showed in figure 6.15, which decreases the time slots utilization efficiency.

Table 6.3: What happened around period 300

Period	296	297	298	299	300	301	302	303
Loss rate in this period %	0.45	0	0	0	99.84	0.10	0.06	0.03
The estimated loss rate for the next period %	0.36 (for 297)	0.05 (for 298)	0.05 (for 299)	0 (for 300)	79.87 (for 301)	10.06 (for 302)	10.05 (for 303)	0.04 (for 304)
Slots assigned to this period	6	5	4	3	2	50	45	46
Slots assigned to the next period	5	4	3	2	50	45	46	45

6.4 Proposal to Solve the Problem

This problem could be avoided if we put a lower and upper bound of the time slots assigned to each period and we can guarantee that the assigned slots are always within the bounds. In this way the case of too few or too many time slots that are going to be assigned to the next period will not happen any more. Then the problem under discussion could be avoided.

For instance, in the example in section 6.3, suppose that we have a lower and upper bound [5, 20]. In period 298, the calculated slots for period 299 is 3 slots which is inferior than the lower bound 5 slots. Then 5 slots, instead of 3 slots, will be assigned to period 299. In the same way, 5 slots, rather than 2 slots, will be assigned to period 300 . By doing so, we will not have a sore of loss rate in period 300. Although the time slots utilization efficiency has decreased a little bit, it is compensated by a lower loss rate and a lower standard deviation.

How to find a functional bound is then the problem that we will face. In this project, we explored Bollinger Bands and Moving Average Envelopes techniques which will be presented and discussed in the following chapters.

Chapter 7

The Simulations with Bounds

As discussed in section 6.3, in the main simulations, the system had poor performances in period 300, especially in the case 1-5, 1-6 and 2-2. In section 6.4, putting bounds to the allocation of each period was proposed so as to avoid this problem. In this chapter, simulations with upper and lower bounds are executed. The results are compared with those of the simulations without the bounds.

We first verified the Bollinger Bands. We found that the idea of bounds help improve the system performance. However, under some situations, the Bollinger Bands do not work well. Thereby, we also tried the Moving Average Envelopes.

7.1 The Definitions of the Bounds

Two types of bounds are defined in this chapter. Both of them are based on the *moving average*, which will be introduced first. Then the two types of bounds will be defined.

7.1.1 The Moving Average

The moving average is an indicator that shows the average value of a data series. To find out the moving average of size 20, we should add up the data values from the past 20^{th} and divide it by 20. Formula 7.1 shows the way to calculate the moving average value with averaged number equal to m.

$$\bar{M}_i = \frac{\sum_{j=i-m+1}^{i} D_j}{m} \tag{7.1}$$

$\bar{M}_i$: the i^{th} moving average value.

m : the number of data being averaged when calculating the moving average. A typical value is 20.

D_j : the j^{th} data value.

7.1.2 The Bollinger Bands (BBs)

The Bollinger Bands consist of three elements: the middle band, the upper band and the lower band. The middle band is the moving average. The upper band is the moving average shifted up by 2 standard deviations. The lower band is the moving average shifted down by 2 standard deviations. They are described by the following formulas.

$$U_i^B = \bar{M}_i + 2\sqrt{\frac{\sum_{j=i-m+1}^{i}(D_j - \bar{M}_i)^2}{m}} \tag{7.2}$$

$$L_i^B = \bar{M}_i - 2\sqrt{\frac{\sum_{j=i-m+1}^{i}(D_j - \bar{M}_i)^2}{m}} \tag{7.3}$$

U_i^B : the i^{th} upper bound value.

L_i^B : the i^{th} lower bound value.

7.1.3 The Moving Average Envelopes (MAEs)

Moving Average Envelopes consists of three bands as well. The middle band is the moving average. The upper band is the moving average shifted up by a percentage amount from a moving average. The lower band is the moving average shifted down by same percentage amount from the moving average. A typical percentage is 20%. They are further described by the following formulas.

$$U_i^B = \bar{M}_i + 0.2\bar{M}_i \tag{7.4}$$

$$L_i^B = \bar{M}_i - 0.2\bar{M}_i \tag{7.5}$$

7.2 Using Upper and Lower Bounds in Our Project

As described in section 6.4, in these simulations the upper and lower bounds will be used in order to improve the performance. The Bollinger Bands and the Moving average envelops will serve respectively as the bounds.

The algorithm is almost the same as the one in section 3.4 except that a comparison with the bounds will be added in the final step. The reader is referred to figure 7.1. At the end of each period, we compare the calculated number of slots with the upper and the lower bound. If the calculated result lies between them, the result is then accepted. Otherwise, if it exceeds the upper bound or the lower bound, then the upper bound or the lower bound will serve as the number of the slots going to be assigned to the next period.

7.3 Simulation Results with the BBs

The results of simulation with Bollinger Bands (BBs) are showed from figure 7.2 to figure 7.12, followed by a comparison between the simulation with and without Bollinger Bands. From these figures, we can conclude:

1. The rise of packet loss rate in period 300 no longer exists in all the cases, as we expected. The time slots utilization efficiency has increased while the bandwidth occupation ratio decreased. Therefore, the performance evaluation

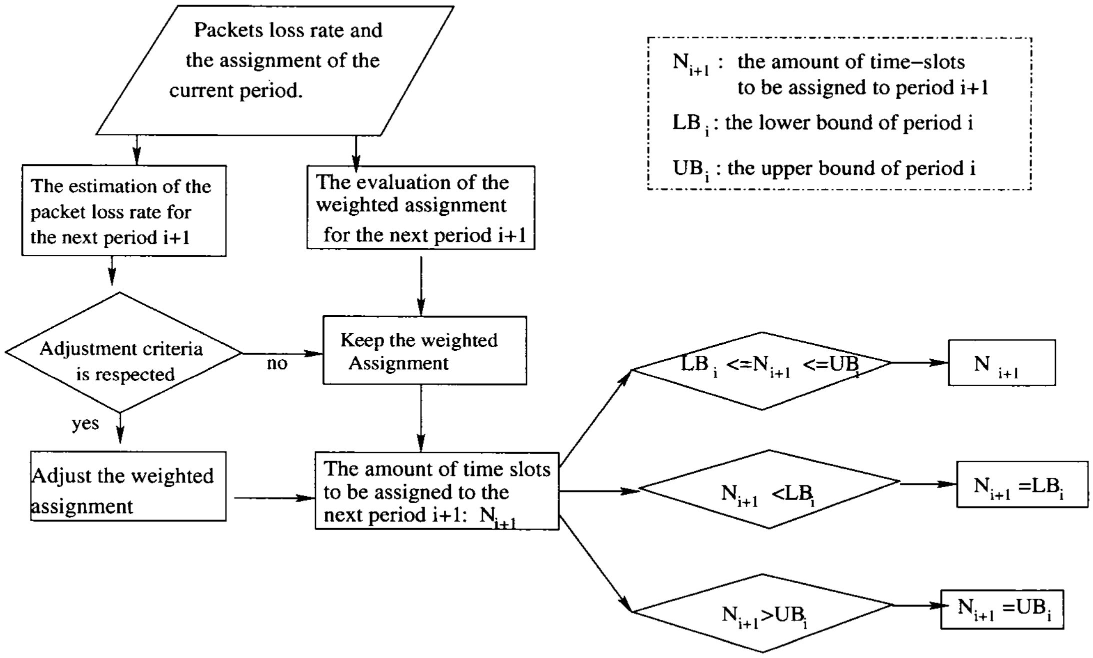

Figure 7.1: Modified algorithm

value has been improved. Especially in case 1-5, 1-6 and 2-2, the improvement is notable. The goal of introducing the BBs is then achieved.

2. Although all the other cases make an improvement, the case 2-6 has poorer performance. The bandwidth utilization ratio decreases from 16.33% to 15.23%. From the figure 7.12, we can notice that the allocation actually converges to 10 time slots and then stays at that value until the last period. This problem is further discussed in section 7.4.

Table 7.1 summarizes the comparison among the cases in the simulation with Bollinger Bands. From this table, we can see that the cases in the 2^{nd} approach work better than those in the 1^{st} approach. The case 2-4 works best. Recall that case 2-6 and 2-4 work best in the simulation without BBs.

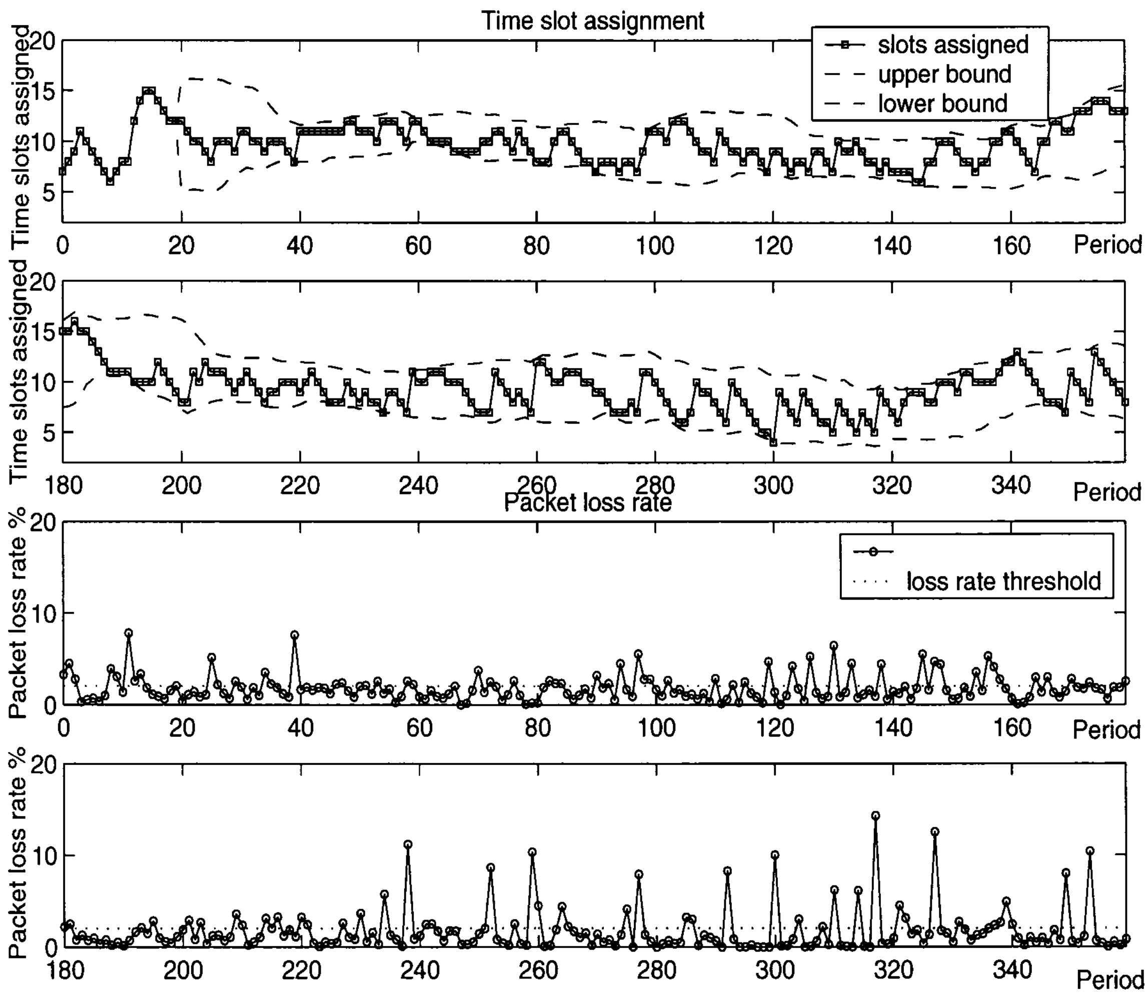

Figure 7.2: Time slots allocation and packet loss rate for case 1-1–BBs

Result comparison of case 1-1

	$\varepsilon\%$	$\bar{R}\ \%$	$\eta\%$	$V\%$	Z
Without	15.23±0.70	1.81±0.2	4.94±0.12	2.38	4.13
With BBs	15.84±0.69	1.78±0.2	4.74±0.10	1.97	4.56

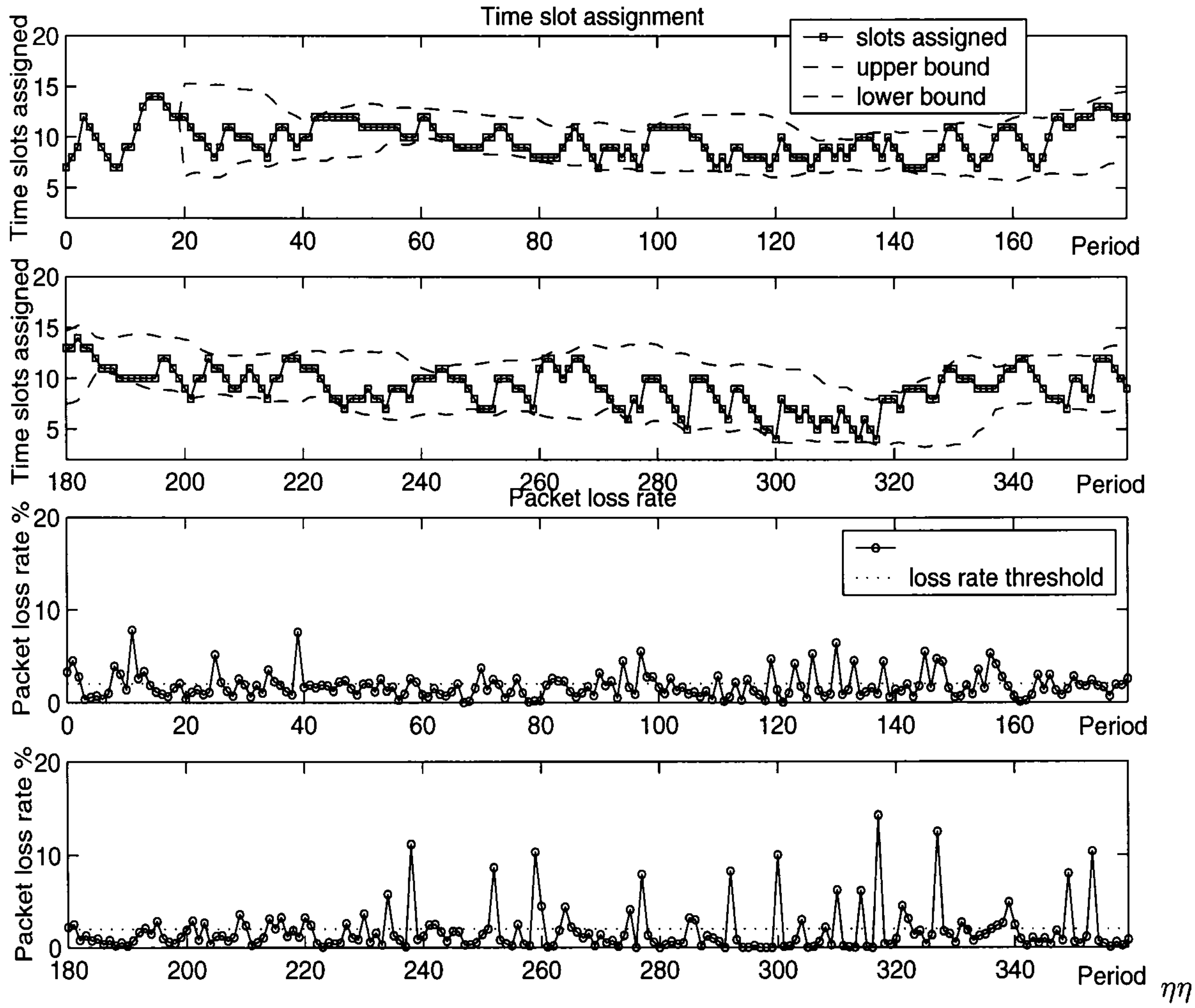

Figure 7.3: Time slots allocation and packet loss rate for case 1-2–BBs

Result comparison of case 1-2

	$\varepsilon\%$	$\bar{R}$ %	$\eta\%$	$V\%$	Z
Without	15.02±0.72	1.76±0.2	5.01±0.12	2.14	4.03
With BBs	15.96±0.70	1.80±0.2	4.71±0.09	2.02	4.56

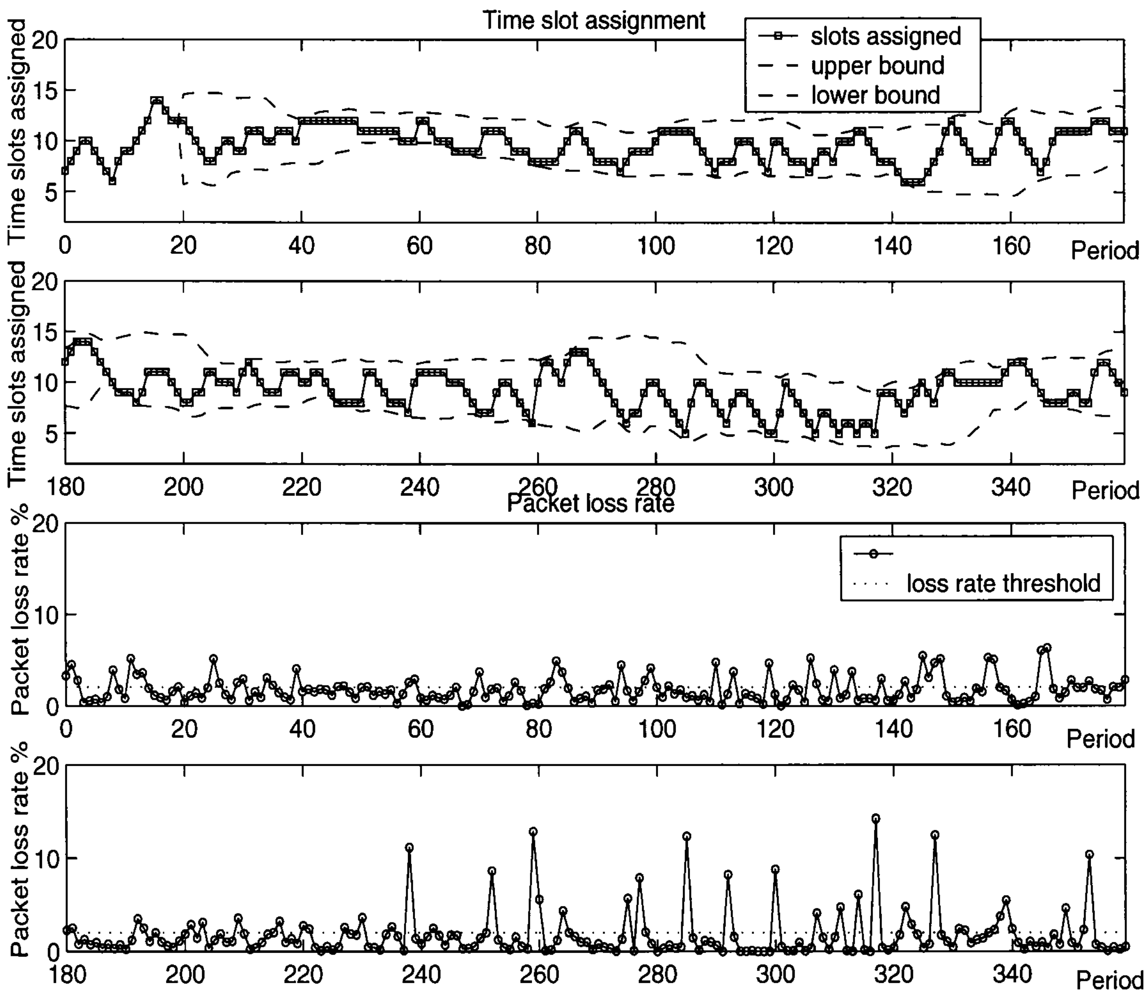

Figure 7.4: Time slots allocation and packet loss rate for case 1-3–BBs

Result comparison of case 1-3

	$\varepsilon\%$	$\bar{R}$ %	$\eta\%$	$V\%$	Z
Without	15.21±0.70	1.77±0.2	4.95±0.12	2.12	4.17
With BBs	15.86±0.70	1.79±0.2	4.74±0.09	2.05	4.51

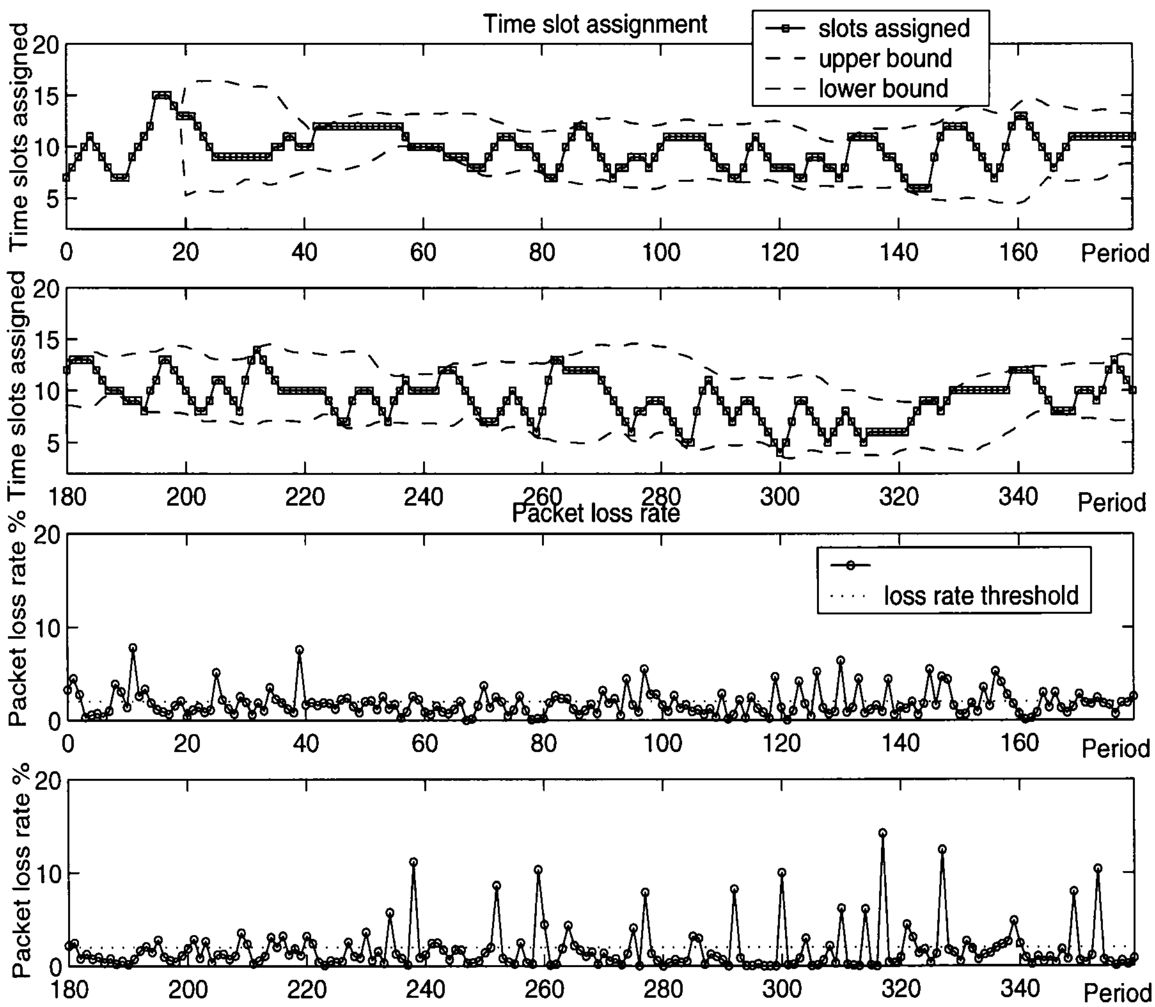

Figure 7.5: Time slots allocation and packet loss rate for case 1-4–BBs

Result comparison of case 1-4

	$\varepsilon\%$	$\bar{R}\,\%$	$\eta\%$	$V\%$	Z
Without	15.14±0.71	1.77±0.2	4.95±0.12	1.95	4.17
With BBs	15.71±0.69	1.78±0.2	4.79±0.10	1.96	4.46

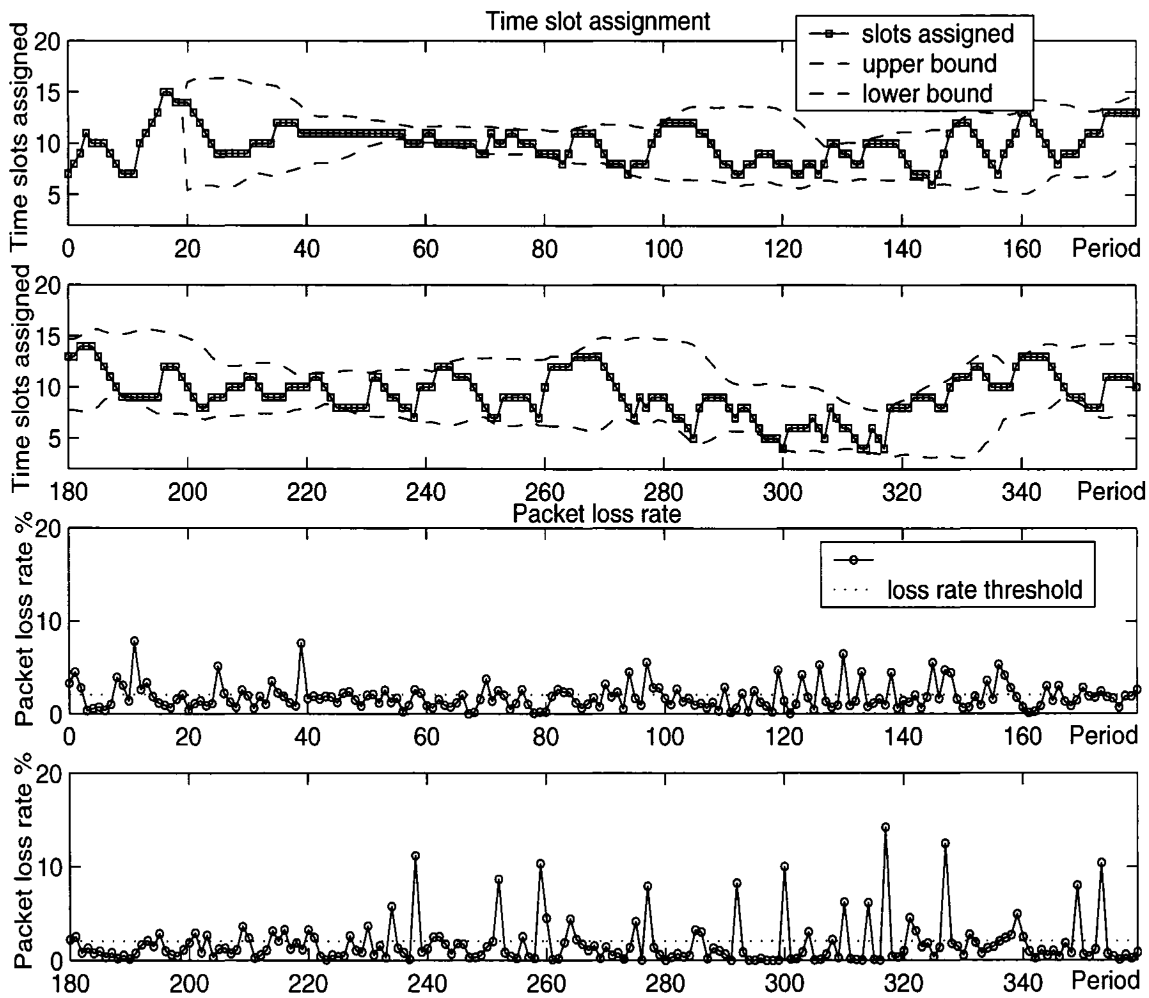

Figure 7.6: Time slots allocation and packet loss rate for case 1-5–BBs

Result comparison of case 1-5

	$\varepsilon\%$	$\bar{R}$ %	$\eta\%$	$V\%$	Z
Without	10.39±0.8	1.83±0.5	7.25±0.60	5.46	0.89
With BBs	15.61±0.70	1.73±0.2	4.82±0.10	2.06	4.83

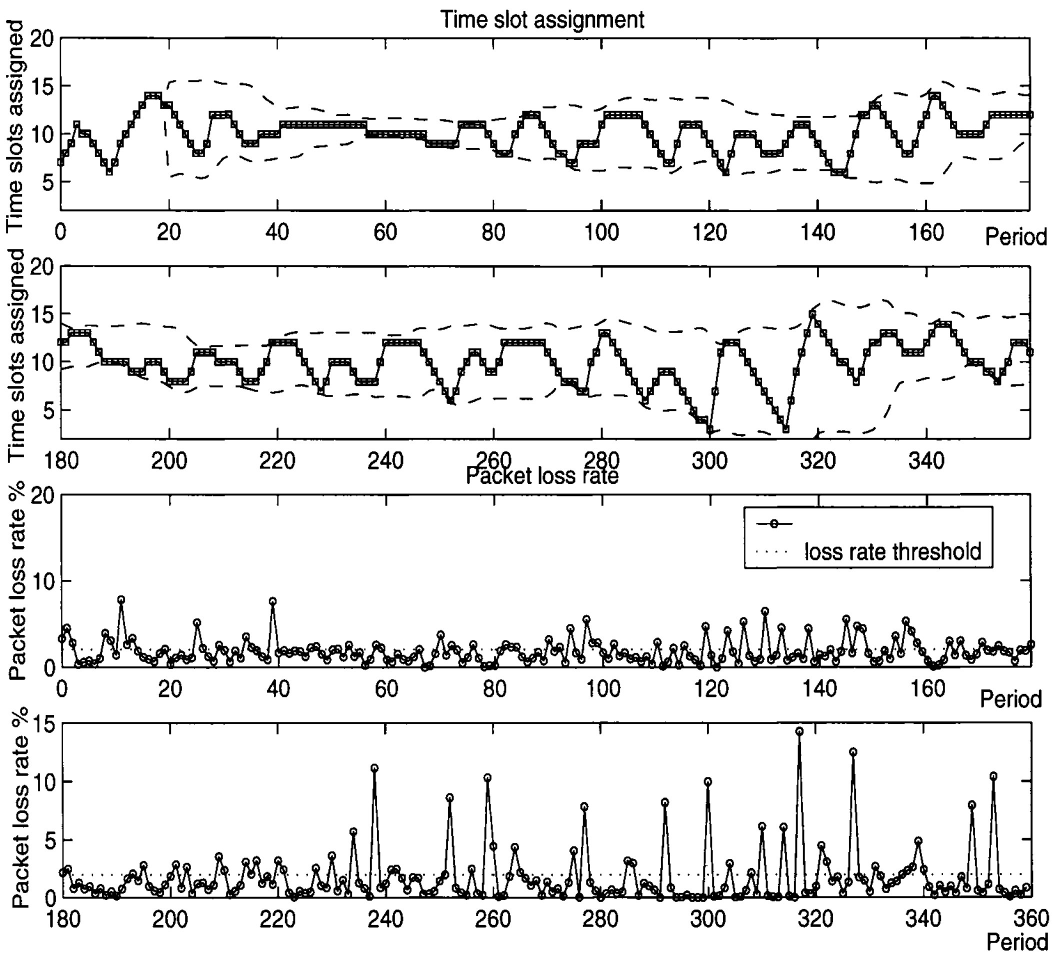

Figure 7.7: Time slots allocation and packet loss rate for case 1-6–BBs

Result comparison of case 1-6

	ε%	$\bar{R}$ %	η%	V%	Z
Without	10.24±0.81	1.86±0.5	7.36±0.58	5.45	0.80
With BBs	15.02±0.72	1.69±0.3	5.02±0.10	2.45	4.01

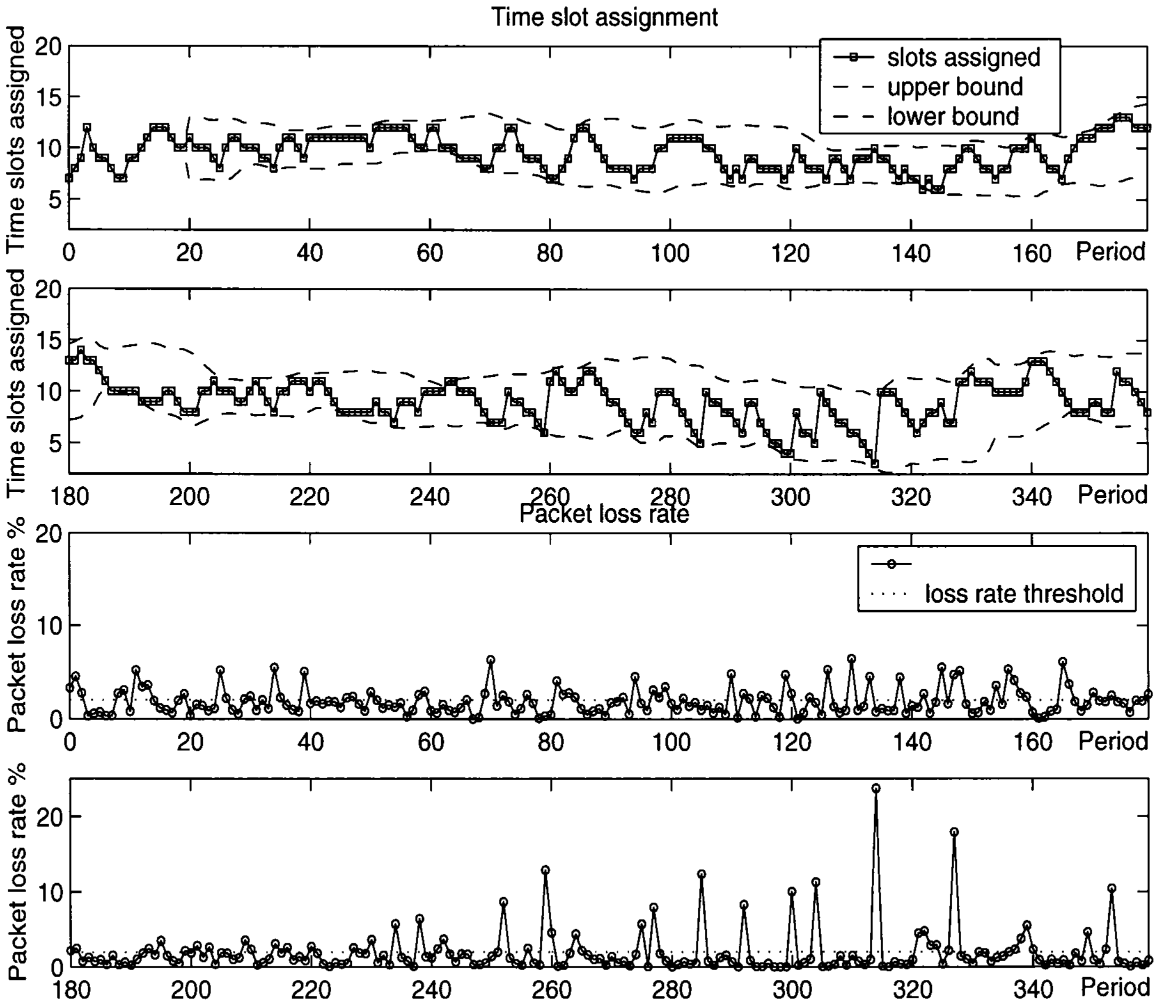

Figure 7.8: Time slots allocation and packet loss rate for case 2-2–BBs

Result comparison of case 2-2

	ε%	$\bar{R}$ %	η%	V%	Z
Without	12.23±0.76	1.93±0.5	6.16±0.40	5.44	1.96
With BBs	16.17±0.70	1.90±0.2	4.64±0.09	2.35	4.61

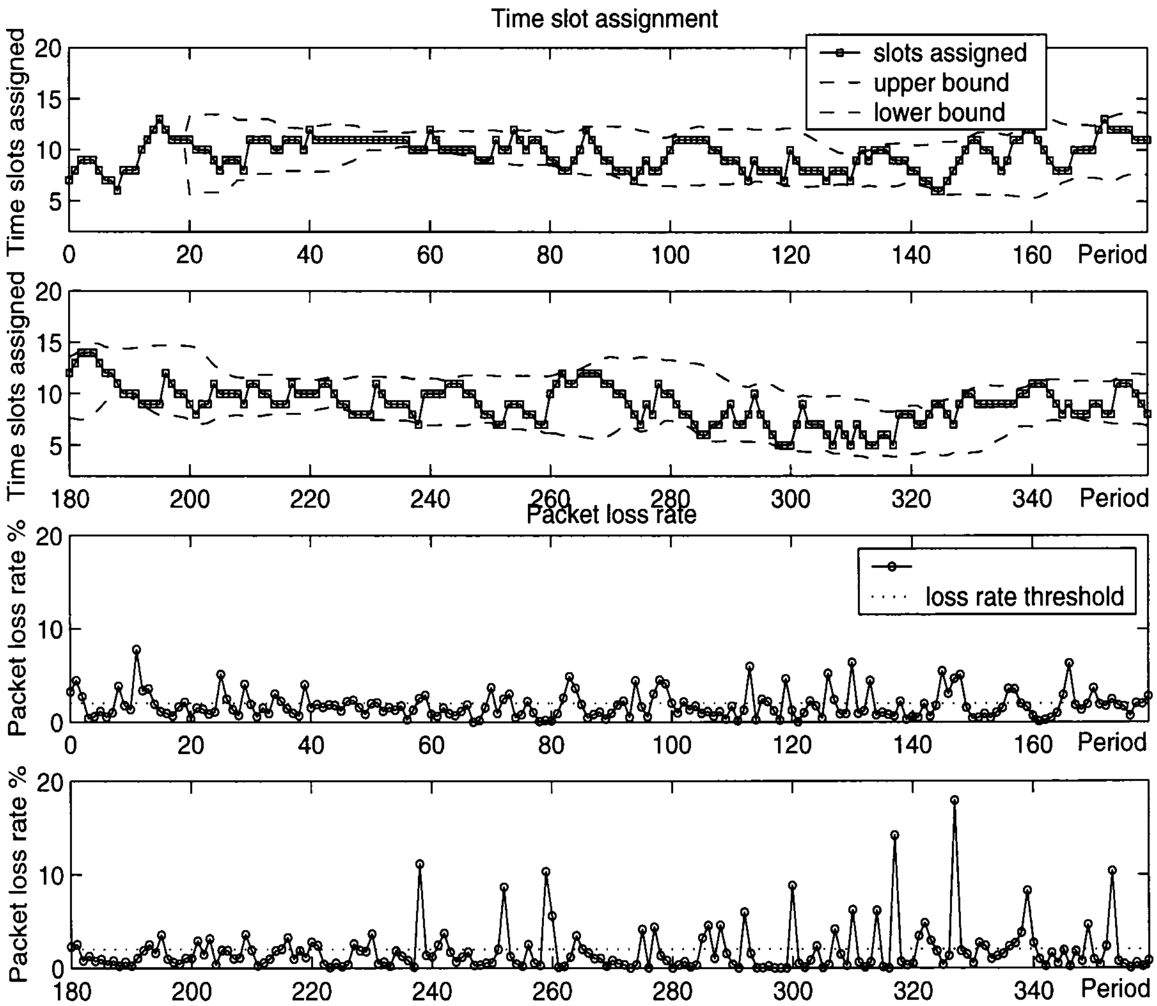

Figure 7.9: Time slots allocation and packet loss rate for case 2-3–BBs

Result comparison of case 2-3

	ε%	$\bar{R}$ %	η%	V%	Z
Without	15.74±0.70	1.79±0.2	4.78±0.09	2.32	4.40
With BBs	16.09±0.70	1.81±0.2	4.67±0.09	2.04	4.63

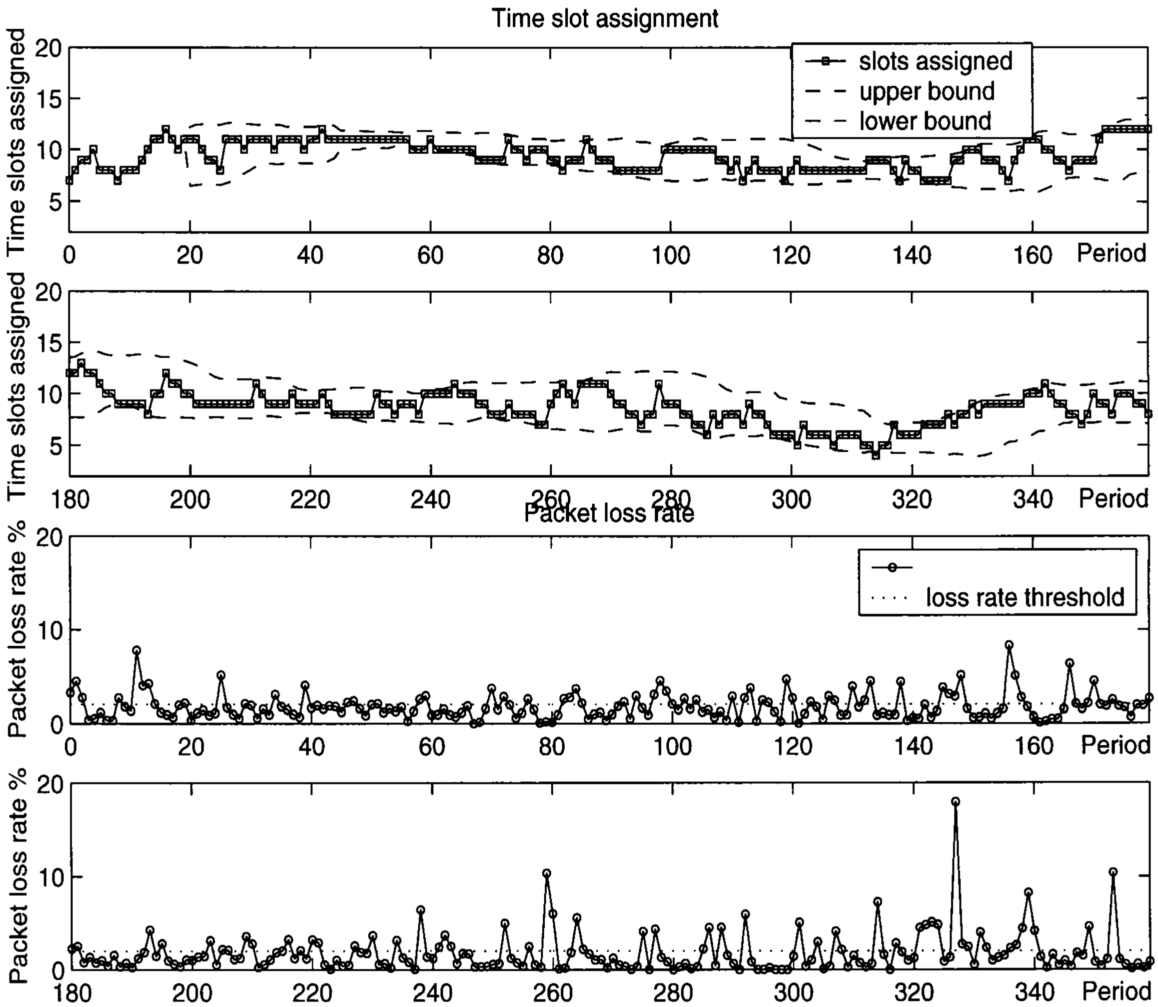

Figure 7.10: Time slots allocation and packet loss rate for case 2-4–BBs

Result comparison of case 2-4

	$\varepsilon\%$	$\bar{R}\ \%$	$\eta\%$	V	Z
Without	16.09±0.7	1.72±0.2	4.68±0.09	1.73	4.69
With BBs	16.64±0.71	1.81±0.2	4.51±0.08	1.84	4.94

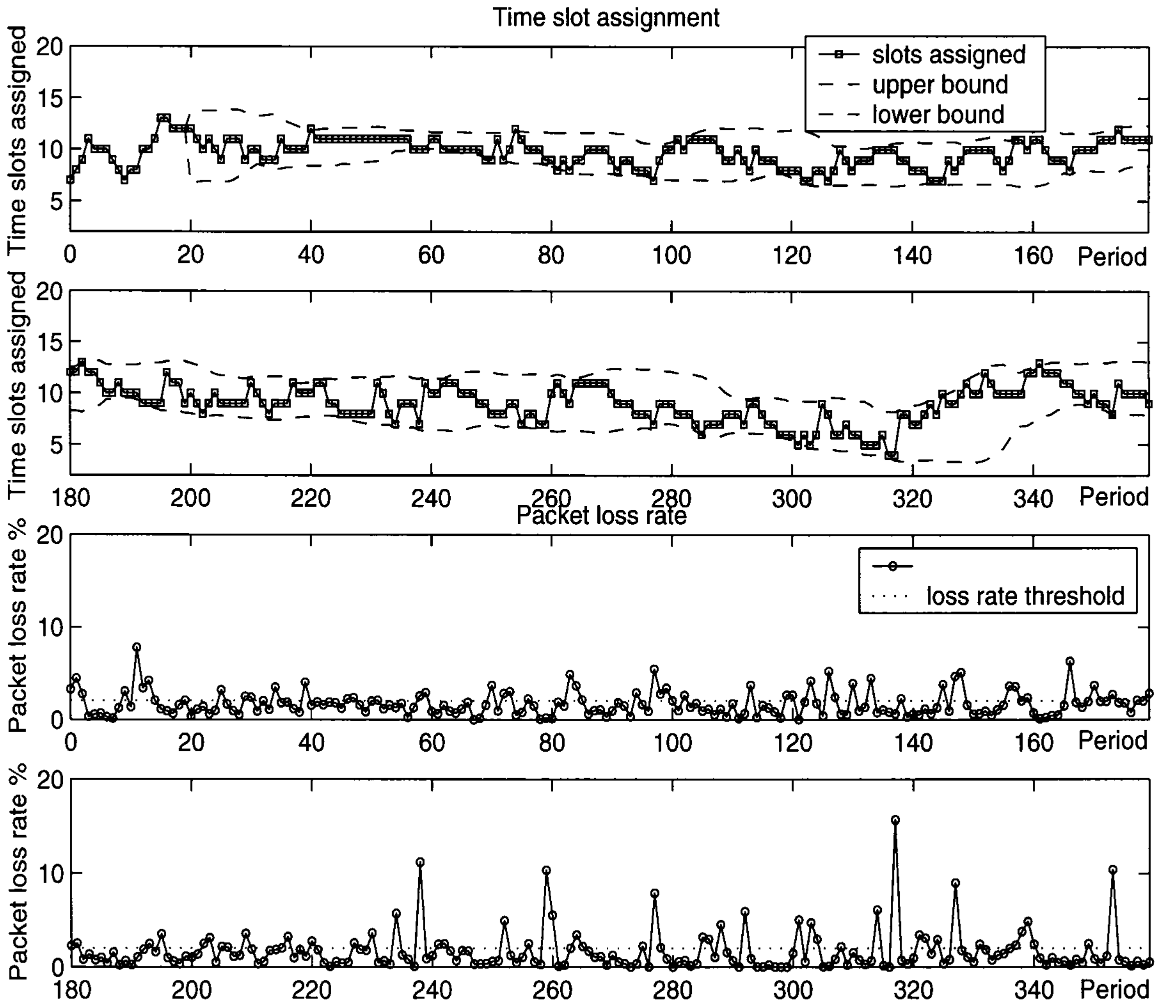

Figure 7.11: Time slots allocation and packet loss rate for case 2-5–BBs

Result comparison of case 2-5

	$\varepsilon\%$	$\bar{R}$ %	$\eta\%$	V	Z
Without	15.75±0.69	1.71±0.2	4.78±0.09	1.94	4.48
With BBs	16.08±0.70	1.69±0.2	4.68±0.08	1.81	4.66

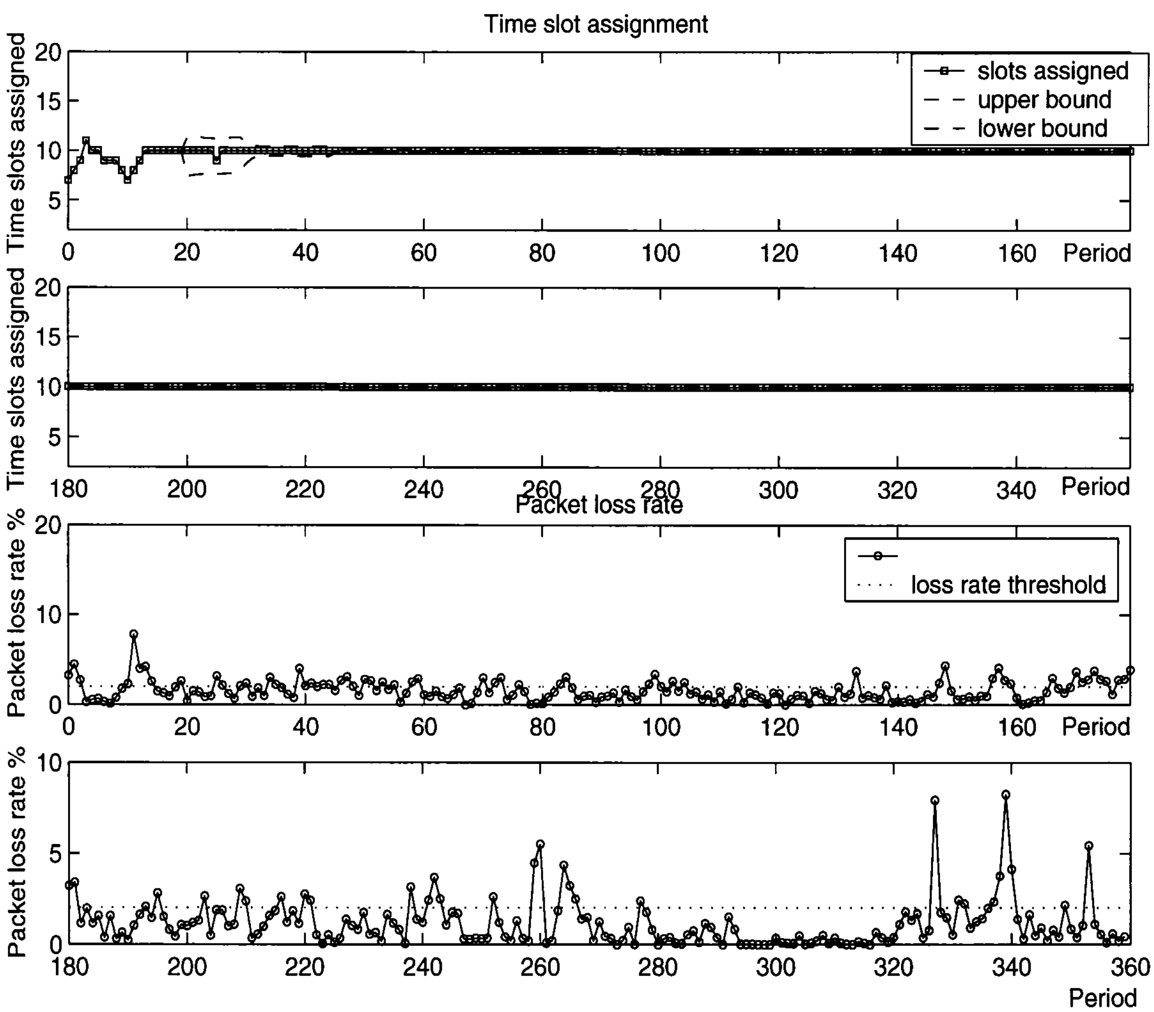

Figure 7.12: Time slots allocation and packet loss rate for case 2-6–BBs

Result comparison of case 2-6

	$\varepsilon\%$	$\bar{R}$ %	$\eta\%$	V	Z
Without	16.33±0.71	1.73±0.2	4.61±0.08	1.69	4.81
With BBs	15.23±0.70	1.39±0.1	4.98±0.02	1.38	4.32

Table 7.1: Performance comparison – simulation with BBs

	Case	ε %	$\bar{R}$ %	η%	V%	Z
1^{st} approach	case 1-1	15.84±0.69	1.78±0.2	4.74±0.10	1.97	4.56
	case 1-2	15.96±0.70	1.80±0.2	4.71±0.09	2.02	4.56
	case 1-3	15.86±0.70	1.79±0.2	4.74±0.09	2.05	4.51
	case 1-4	15.71±0.69	1.78±0.2	4.79±0.10	1.96	4.46
	case 1-5	15.61±0.70	1.73±0.2	4.82±0.10	2.06	4.83
	case 1-6	15.02±0.72	1.69±0.3	5.02±0.10	2.45	4.01
2^{nd} approach	case 2-2	16.17±0.70	1.90±0.2	4.64±0.09	2.35	4.61
	case 2-3	16.09±0.70	1.81±0.2	4.67±0.09	2.04	4.63
	case 2-4	16.64±0.71	1.81±0.2	4.51±0.08	1.84	4.94
	case 2-5	16.08±0.70	1.69±0.2	4.68±0.08	1.81	4.66
	case 2-6	15.23±0.70	1.39±0.1	4.98±0.02	1.38	4.32

Figure 7.13 shows the histogram of time slots assigned to each period. Comparing this figure with the histogram of the simulation without Bollinger Bands (see figure 6.12), we can see that the percentages of the cases in which more than 15 slots are assigned to a period decreases. This agrees to the conclusion that the bandwidth occupation ratio is decreased compared with the simulation without BBs.

7.4 The Problem with the BBs

We notice that in case 2-6 the time slots allocation converges to 10 slots and stays fixed to this value until the last period without any change. This phenomenon can be explained from the formula of the Bollinger Bands. Recall the formula 7.2,

$$U_i^B = \bar{M}_i + 2\sqrt{\frac{\sum_{j=i-m+1}^{i}(D_j - \bar{M}_i)^2}{m}}$$

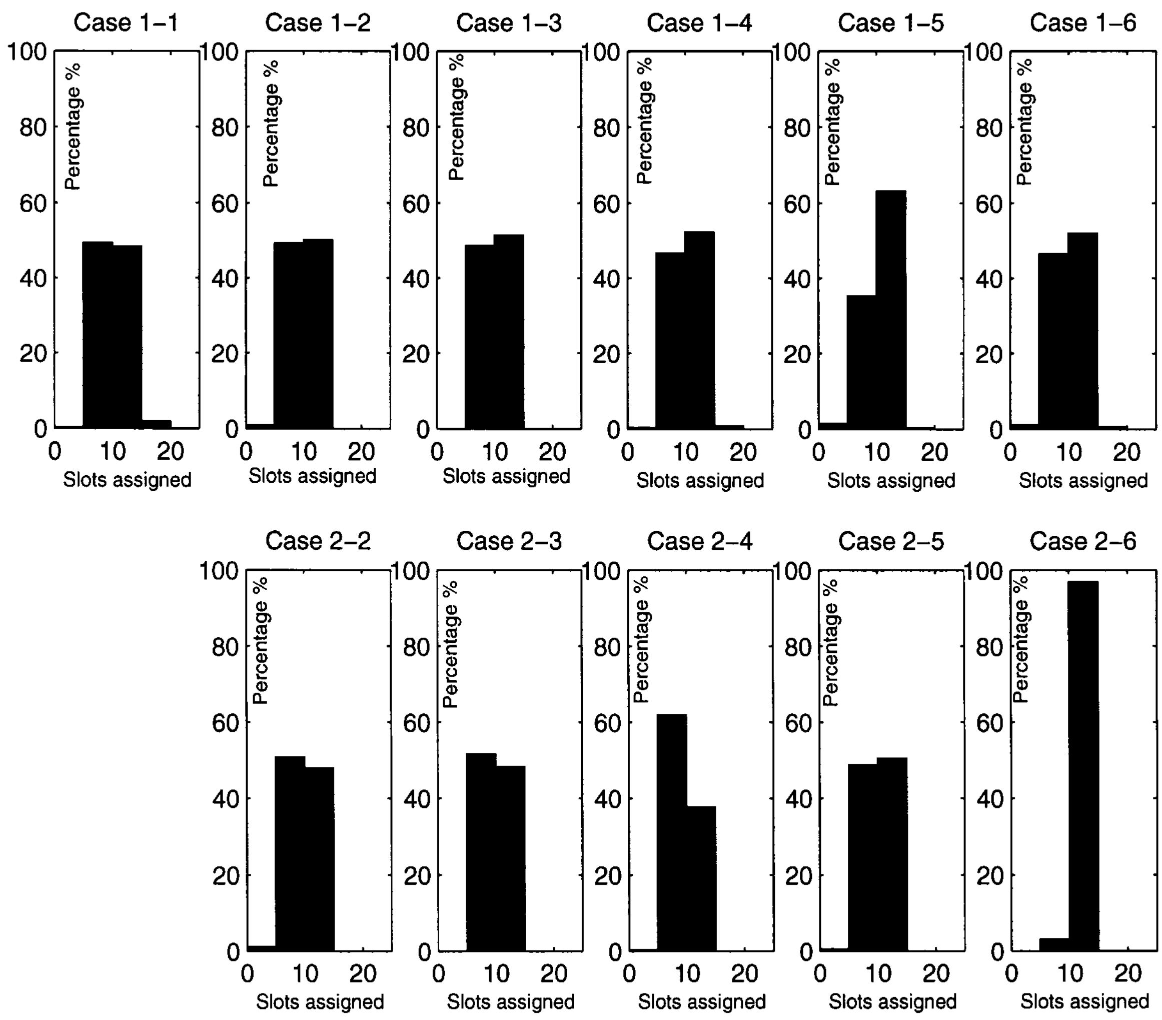

Figure 7.13: The histogram of time slots assigned—simulation with BBs

$$L_i^B = \bar{M}_i - 2\sqrt{\frac{\sum_{j=i-m+1}^{i}(D_j - \bar{M}_i)^2}{m}}$$

if, for some i, we happen to have:

$$D_j = \bar{M}_i, \text{ for } j = i - m + 1 \text{ to } i$$

then we will have

$$U_i^B = \bar{M}_i = L_i^B,$$

which means the upper bound, the lower bound and the moving average are of the same value. Since the allocation should be within the bounds, we will then have an allocation equal to the bounds value. Furthermore, this allocation can not be adjusted.

That explains the convergence happening in the case 2-6. If this happens, the allocation can not be adjusted according to the performance. It will decrease the time slots utilization efficiency and the performance evaluation value.

From the definition of the Moving average envelopes, we can see that the envelopes are shifted by a fixed percentage from the moving average. Thus, the convergence problem will not happen in any situation. It is then interesting to use the Moving average envelopes as the upper and lower bounds.

7.5 Simulation Results with the MAEs

The same simulation is done with Moving Average Envelopes. The results are shown from figure 7.14 to figure 7.24, followed by a comparison between the simulation with and without MAEs.

From these figures, we conclude that:

1. The rise of packet loss rate in period 300 does not exist either. Without the rise in period 300, the performance of cases 1-5, 1-6 and 2-2 has been greatly improved.

2. The convergence phenomenon that happened in simulation with BBs does not occur here, as we expected. This means that MAEs can correct the problem brought by BBs.

3. Moreover, the objective function value has been further increased in all the cases compared with those with BBs.

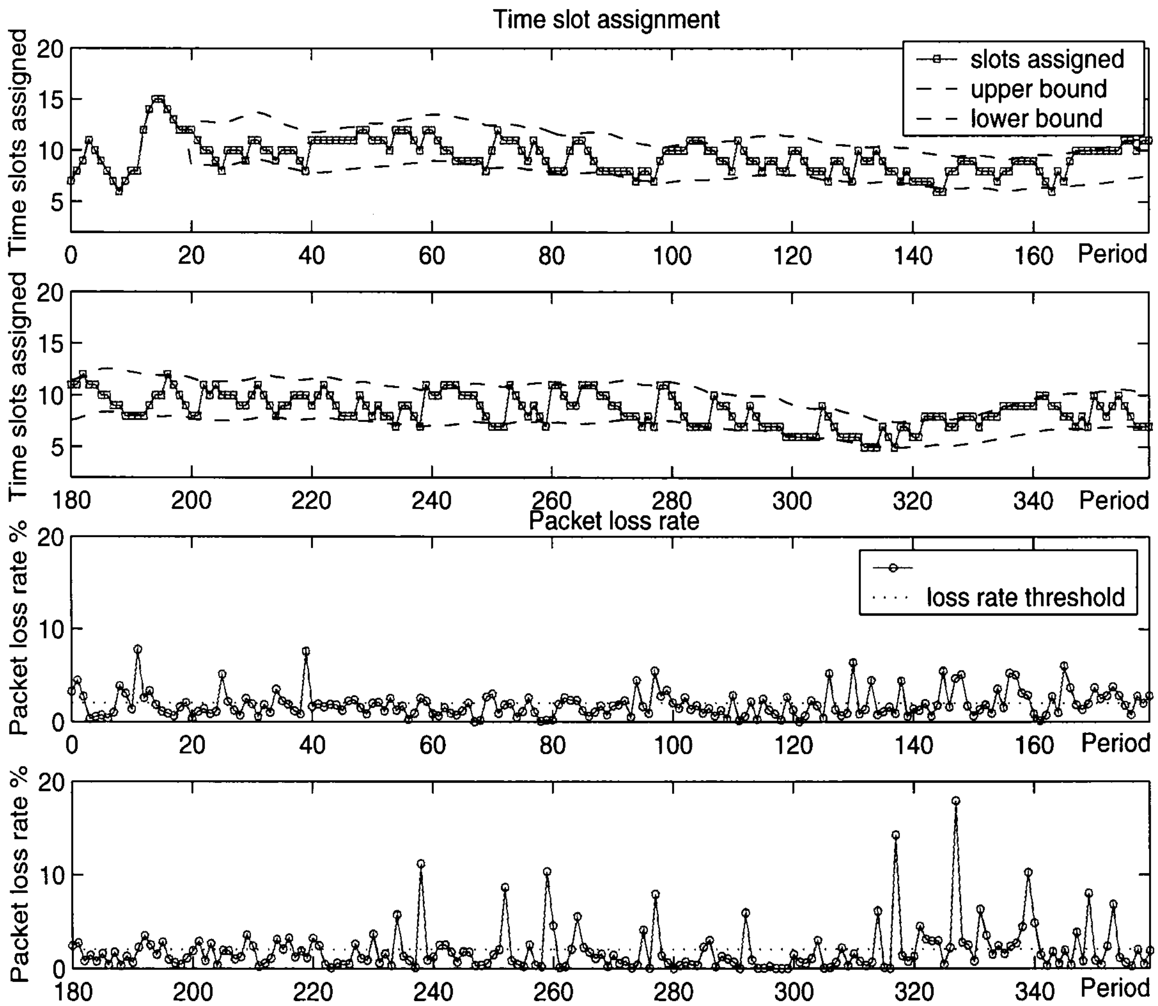

Figure 7.14: Time slots allocation and packet loss rate for case 1-1—MAEs

Result comparison of case 1-1

	ε %	$\bar{R}$ %	η%	V%	Z
Without	15.23±0.70	1.81±0.2	4.94±0.12	2.38	4.13
With MAEs	16.62±0.71	1.91±0.2	4.51±0.08	2.04	4.88

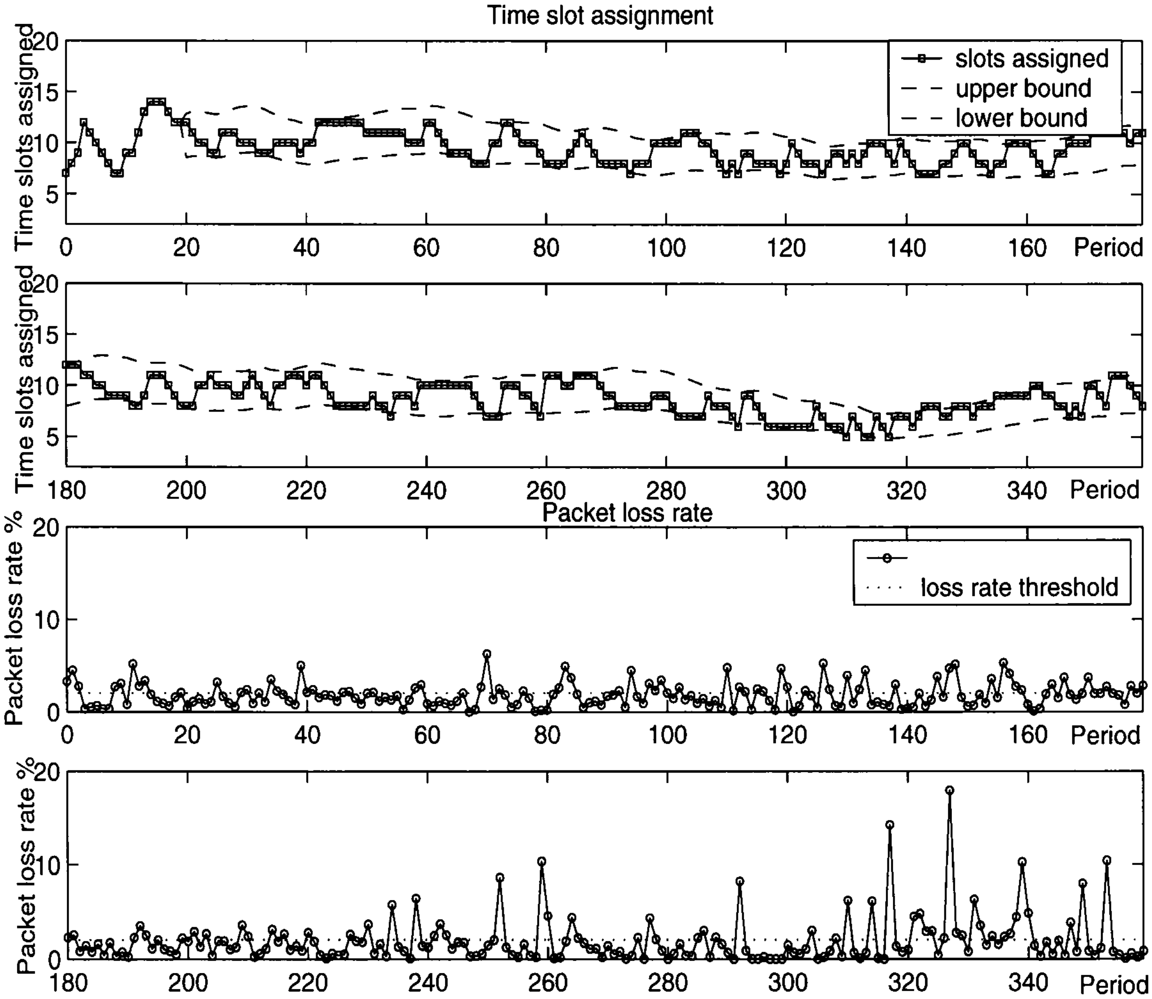

Figure 7.15: Time slots allocation and packet loss rate for case 1-2—MAEs

Result comparison of case 1-2

	ε %	$\bar{R}$ %	η%	V%	Z
Without	15.02±0.72	1.76±0.2	5.01±0.12	2.14	4.03
With MAEs	16.49±0.70	1.87±0.2	4.55±0.08	2.02	4.83

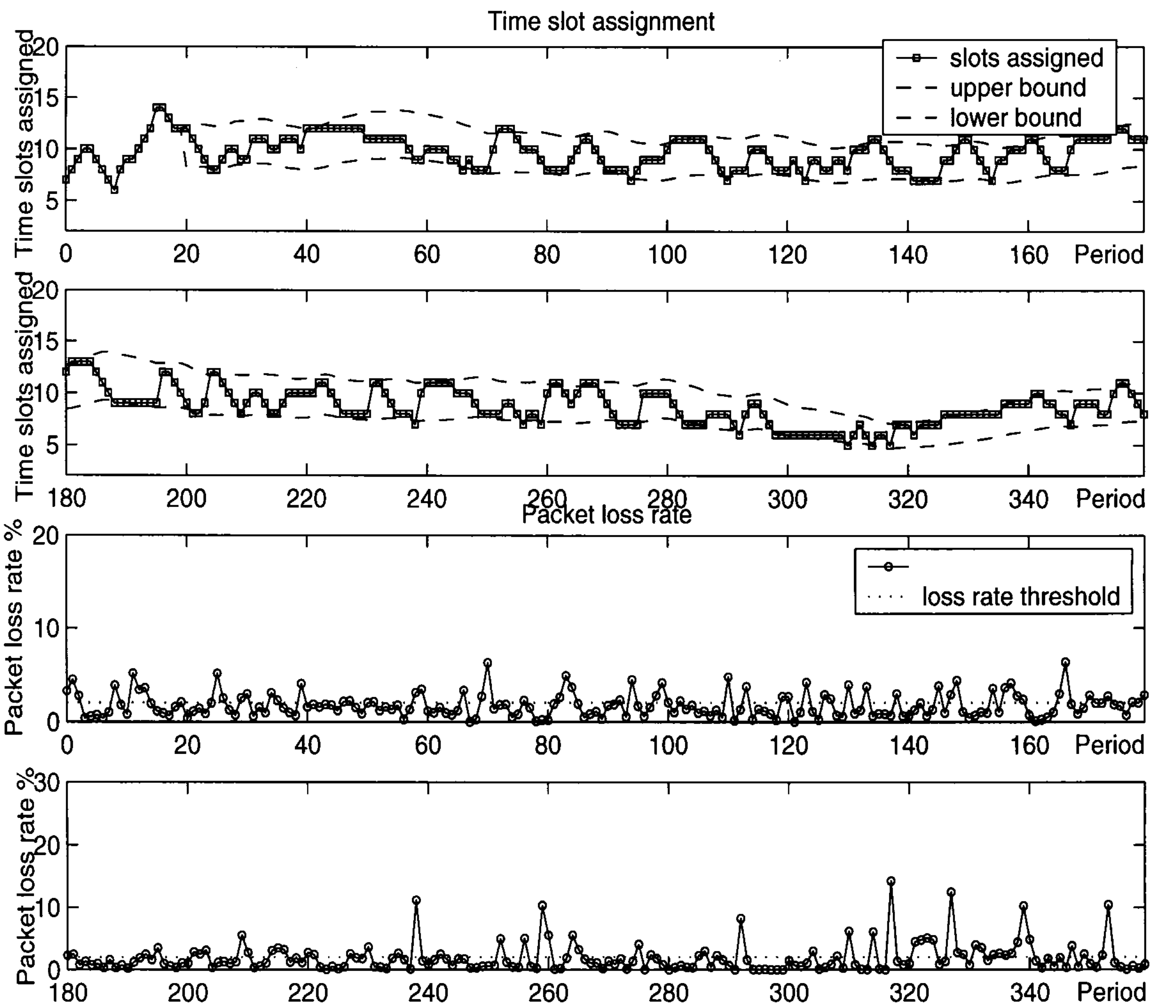

Figure 7.16: Time slots allocation and packet loss rate for case 1-3—MAEs

Result comparison of case 1-3

	ε %	$\bar{R}$ %	η%	V%	Z
Without	15.21±0.70	1.77±0.2	4.95±0.12	2.12	4.17
With MAEs	16.31±0.71	1.82±0.2	4.60±0.09	1.89	4.77

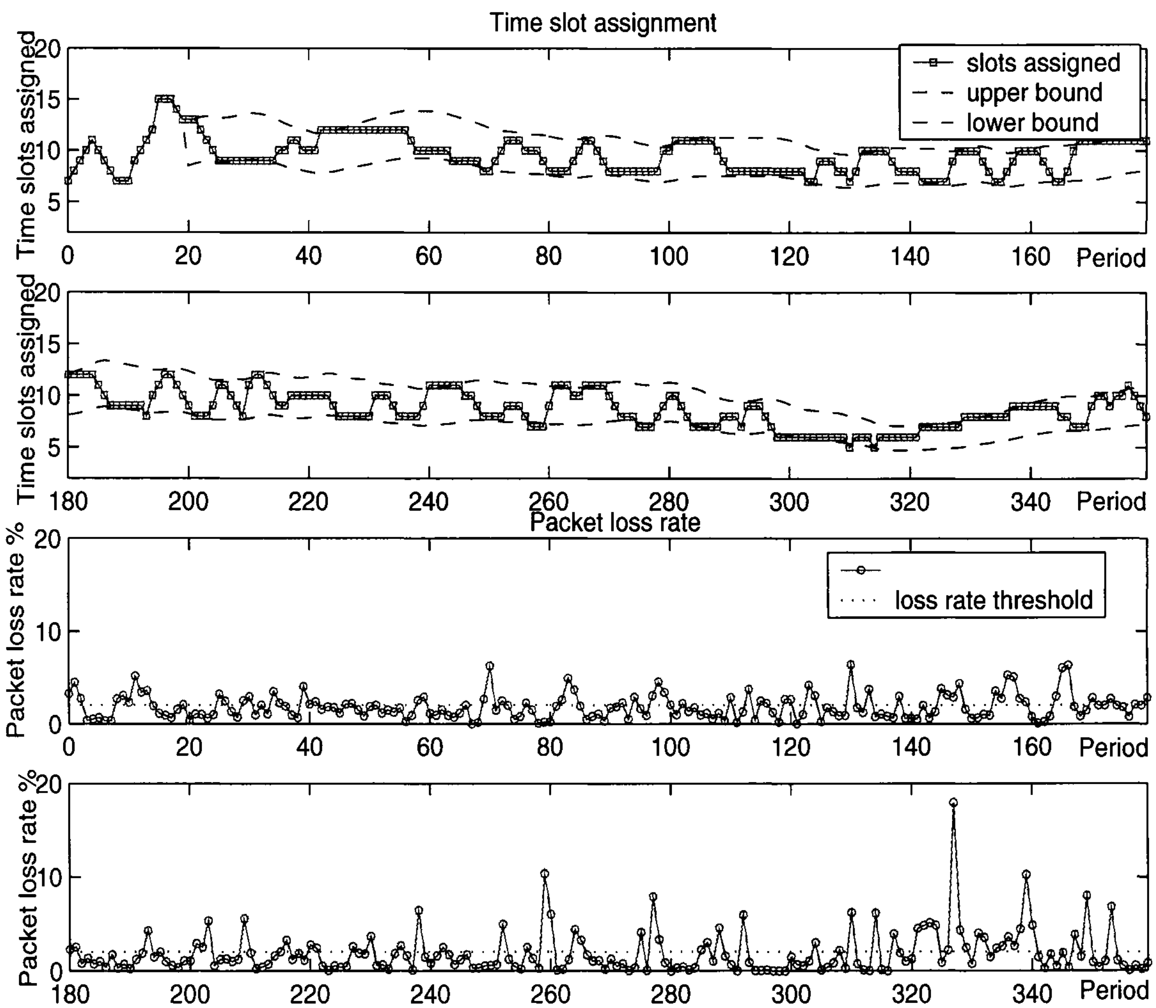

Figure 7.17: Time slots allocation and packet loss rate for case 1-4—MAEs

Result comparison of case 1-4

	ε %	$\bar{R}$ %	η%	V%	Z
Without	15.14±0.71	1.77±0.2	4.95±0.12	1.95	4.17
With MAEs	16.44±0.70	1.86±0.2	4.57±0.08	1.87	4.83

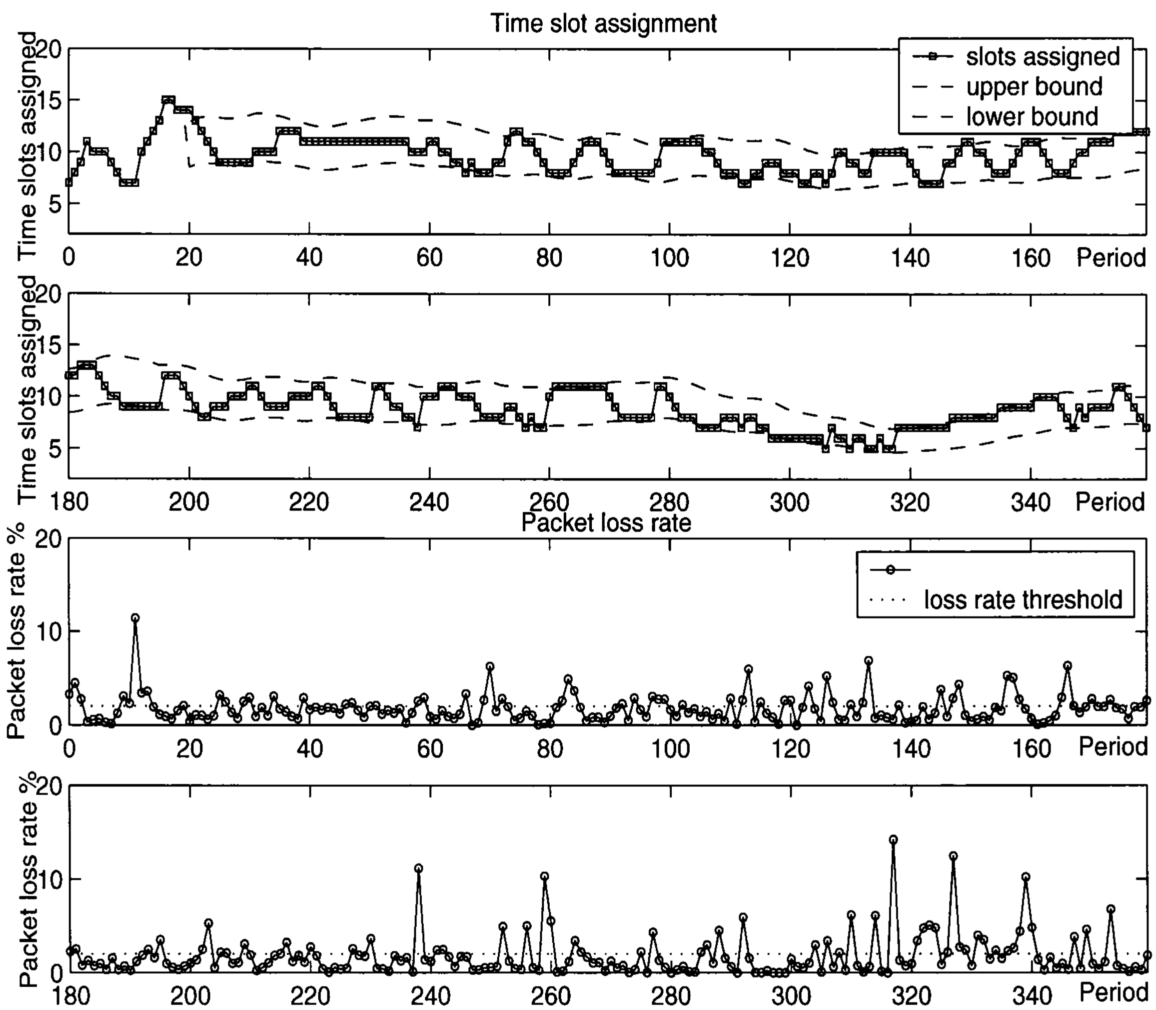

Figure 7.18: Time slots allocation and packet loss rate for case 1-5—MAEs

Result comparison of case 1-5

	ε %	$\bar{R}$ %	η%	V%	Z
Without	10.39±0.8	1.83±0.5	7.25±0.60	5.46	0.89
With MAEs	16.24±0.71	1.81±0.2	4.63±0.09	1.90	4.73

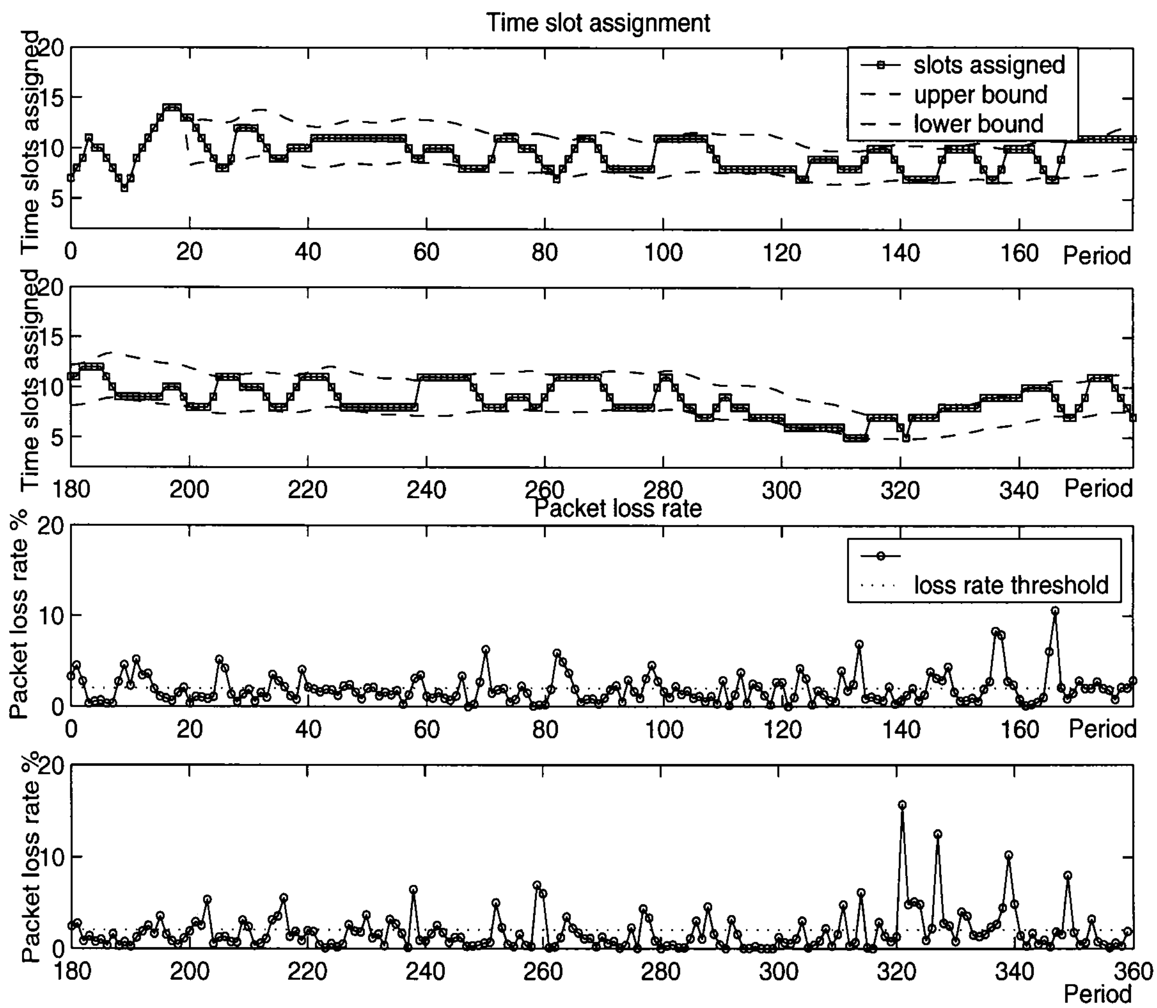

Figure 7.19: Time slots allocation and packet loss rate for case 1-6—MAEs

Result comparison of case 1-6

	ε %	$\bar{R}$ %	η%	V%	Z
Without	10.24±0.81	1.86±0.5	7.36±0.58	5.45	0.80
With MAEs	16.40±0.70	1.85±0.2	4.58±0.08	1.90	4.80

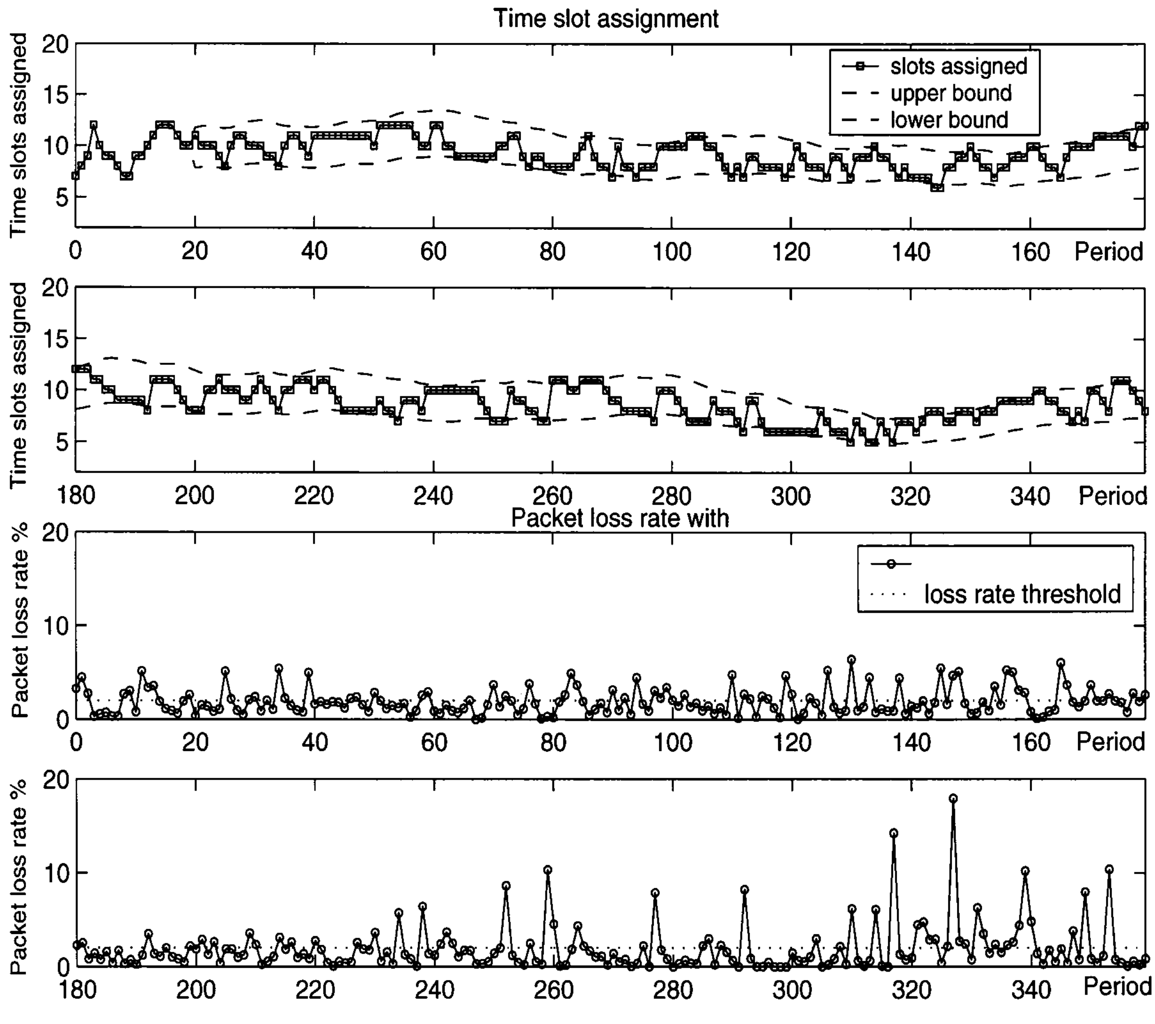

Figure 7.20: Time slots allocation and packet loss rate for case 2-2—MAEs

Result comparison of case 2-2

	ε %	$\bar{R}$ %	η%	V%	Z
Without	12.23±0.76	1.93±0.5	6.16±0.40	5.44	1.96
With MAEs	16.67±0.70	1.92±0.2	4.50±0.08	2.04	4.91

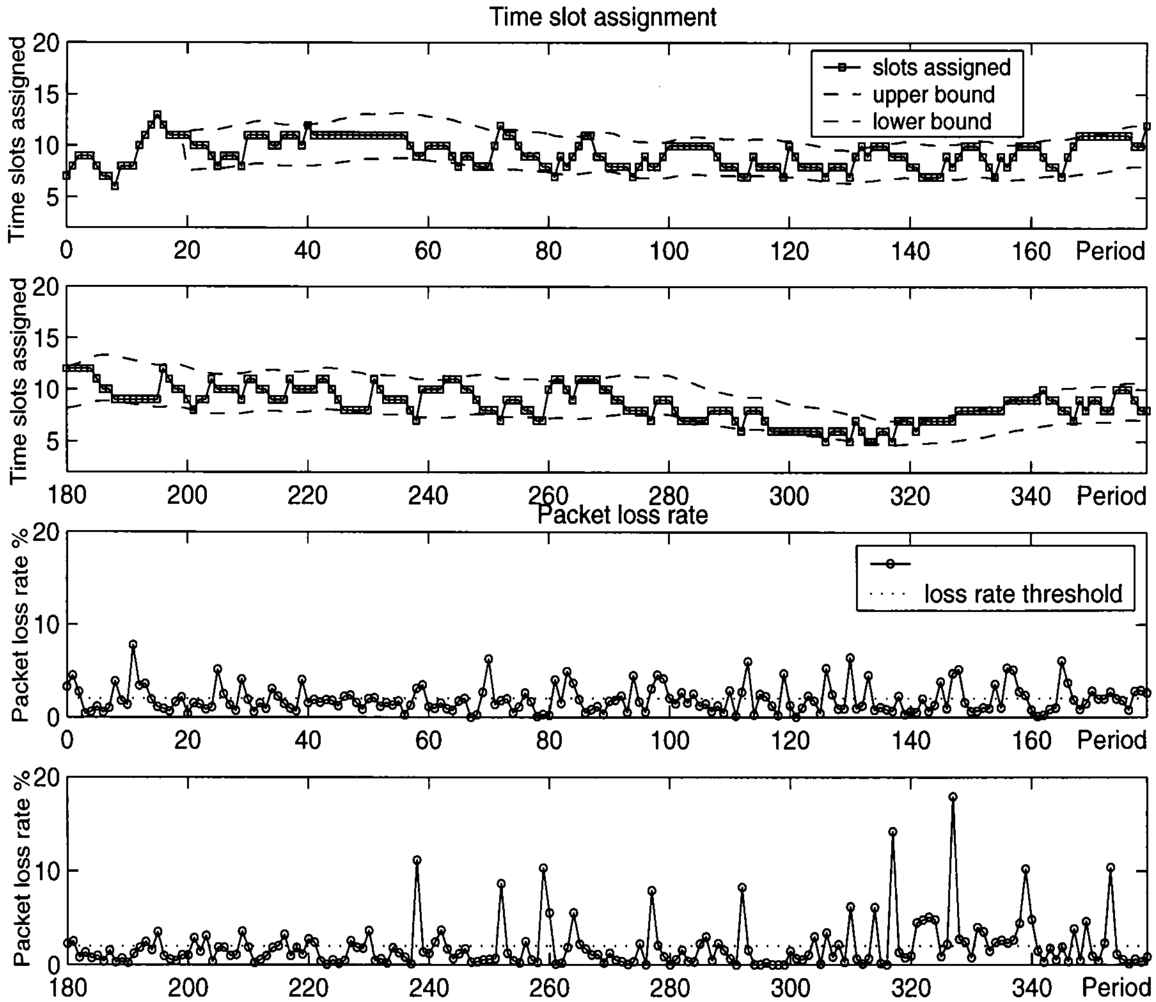

Figure 7.21: Time slots allocation and packet loss rate for case 2-3—MAEs

Result comparison of case 2-3

	ε %	$\bar{R}$ %	η%	V%	Z
Without	15.74±0.70	1.79±0.2	4.78±0.09	2.32	4.40
With MAEs	16.70±0.71	1.93±0.2	4.49±0.08	2.09	4.92

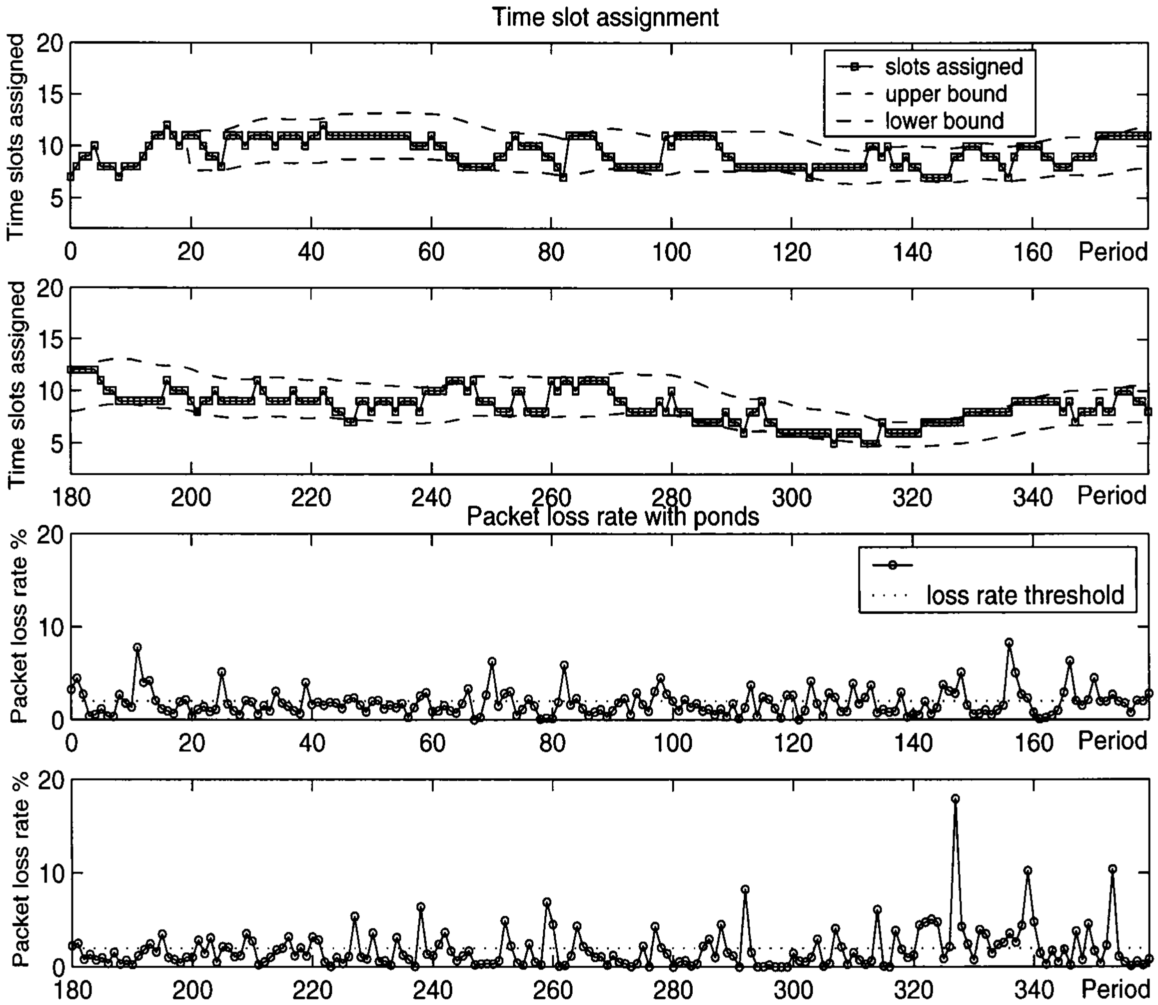

Figure 7.22: Time slots allocation and packet loss rate for case 2-4—MAEs

Result comparison of case 2-4

	ε %	$\bar{R}$ %	η%	V%	Z
Without	16.09±0.7	1.72±0.2	4.68±0.09	1.73	4.69
With MAEs	16.74±0.71	1.84±0.2	4.49±0.08	1.84	4.92

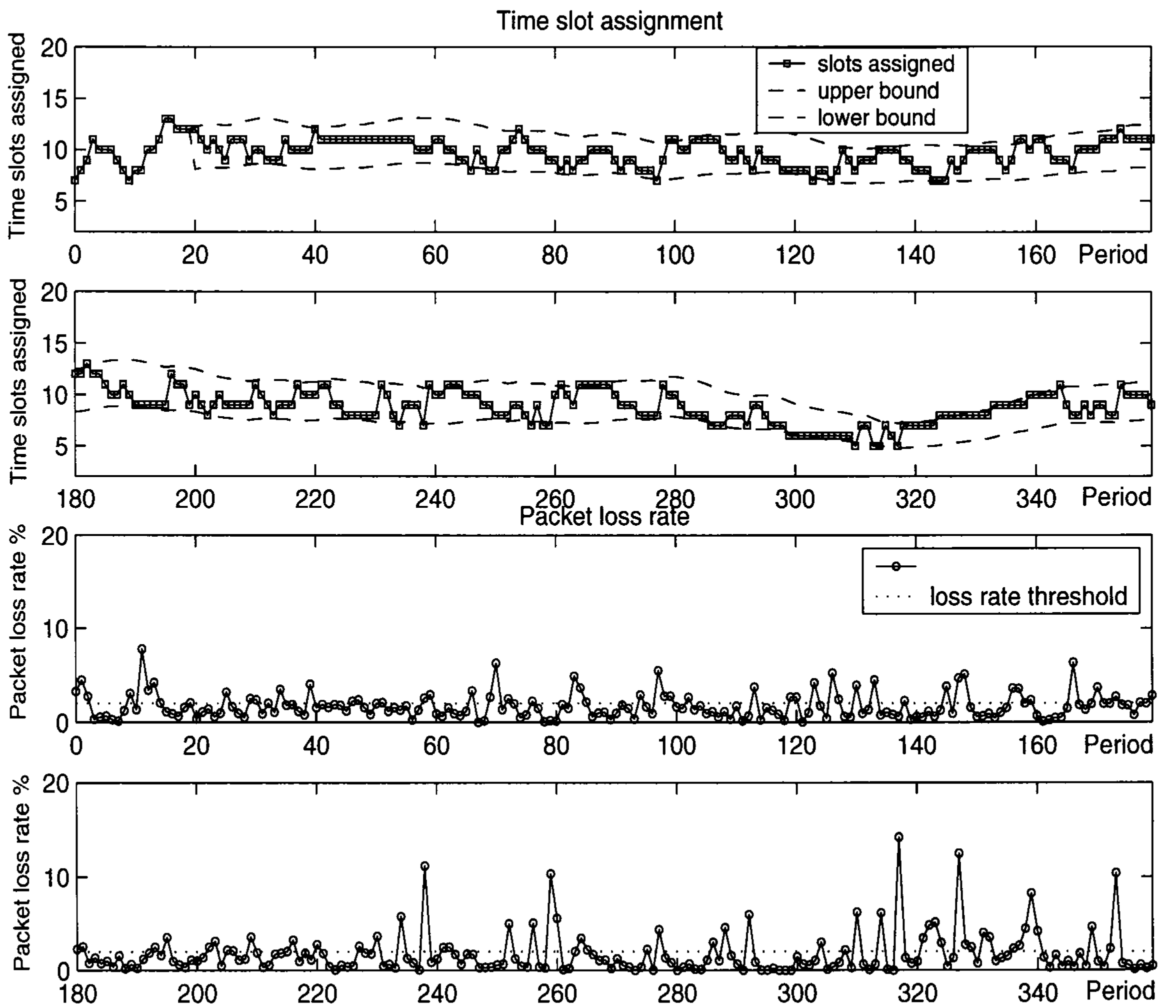

Figure 7.23: Time slots allocation and packet loss rate for case 2-5—MAEs

Result comparison of case 2-5

	ε %	$\bar{R}$ %	η%	V%	Z
Without	15.75±0.69	1.71±0.2	4.78±0.09	1.94	4.48
With MAEs	16.27±0.70	1.76±0.2	4.62±0.08	1.88	4.75

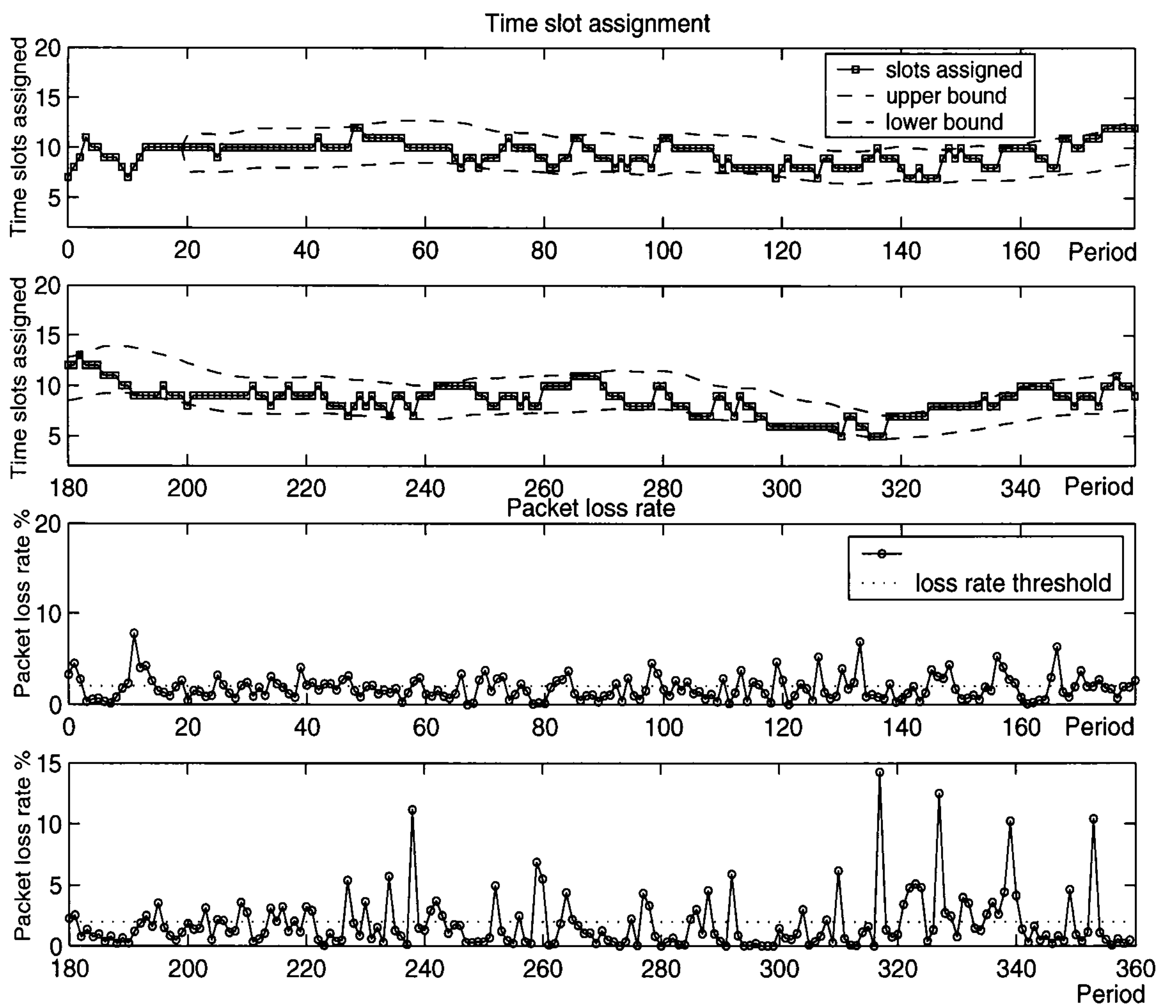

Figure 7.24: Time slots allocation and packet loss rate for case 2-6—MAEs

Result comparison of case 2-6

	ε %	$\bar{R}$ %	η%	V%	Z
Without	16.33±0.71	1.73±0.2	4.61±0.08	1.69	4.81
With MAE	16.63±0.72	1.83±0.2	4.52±0.07	1.86	4.93

Table 7.2 summarizes the comparison among the cases in the simulation with the Moving Average Envelopes. We can see that the cases in the 2^{nd} approach present a better performance than those in the 1^{st} approach. The case 2-4 performs best.

Figure 7.25 shows the histogram of time slots assigned to each period. Comparing this figure with the histogram of the simulation without MAEs (see figure 6.12), we can see that the percentages of the times in which more than 15 slots are assigned to a period decrease, while the percentages of the times in which less than 10 time slots are assigned increase. This agrees to the conclusions we had drawn.

Table 7.2: Performance comparison – simulation with MAEs

	Case	ε %	$\bar{R}$ %	η%	V%	Z
1^{st} approach	case 1-1	16.62±0.71	1.91±0.2	4.51±0.08	2.04	4.88
	case 1-2	16.49±0.70	1.87±0.2	4.55±0.08	2.02	4.83
	case 1-3	16.31±0.71	1.82±0.2	4.60±0.09	1.89	4.77
	case 1-4	16.44±0.70	1.86±0.2	4.57±0.08	1.87	4.83
	case 1-5	16.24±0.71	1.81±0.2	4.63±0.09	1.90	4.73
	case 1-6	16.40±0.70	1.85±0.2	4.58±0.08	1.90	4.80
2^{nd} approach	case 2-2	16.67±0.70	1.92±0.2	4.50±0.08	2.04	4.91
	case 2-3	16.70±0.71	1.93±0.2	4.49±0.08	2.09	4.92
	case 2-4	16.74±0.71	1.84±0.2	4.49±0.08	1.84	4.92
	case 2-5	16.27±0.70	1.76±0.2	4.62±0.08	1.88	4.75
	case 2-6	16.63±0.72	1.83±0.2	4.52±0.07	1.86	4.93

7.6 Summary and Conclusions

The results of the main simulation and the simulations with bounds are summarized in table 7.3, according to which we conclude as follows.

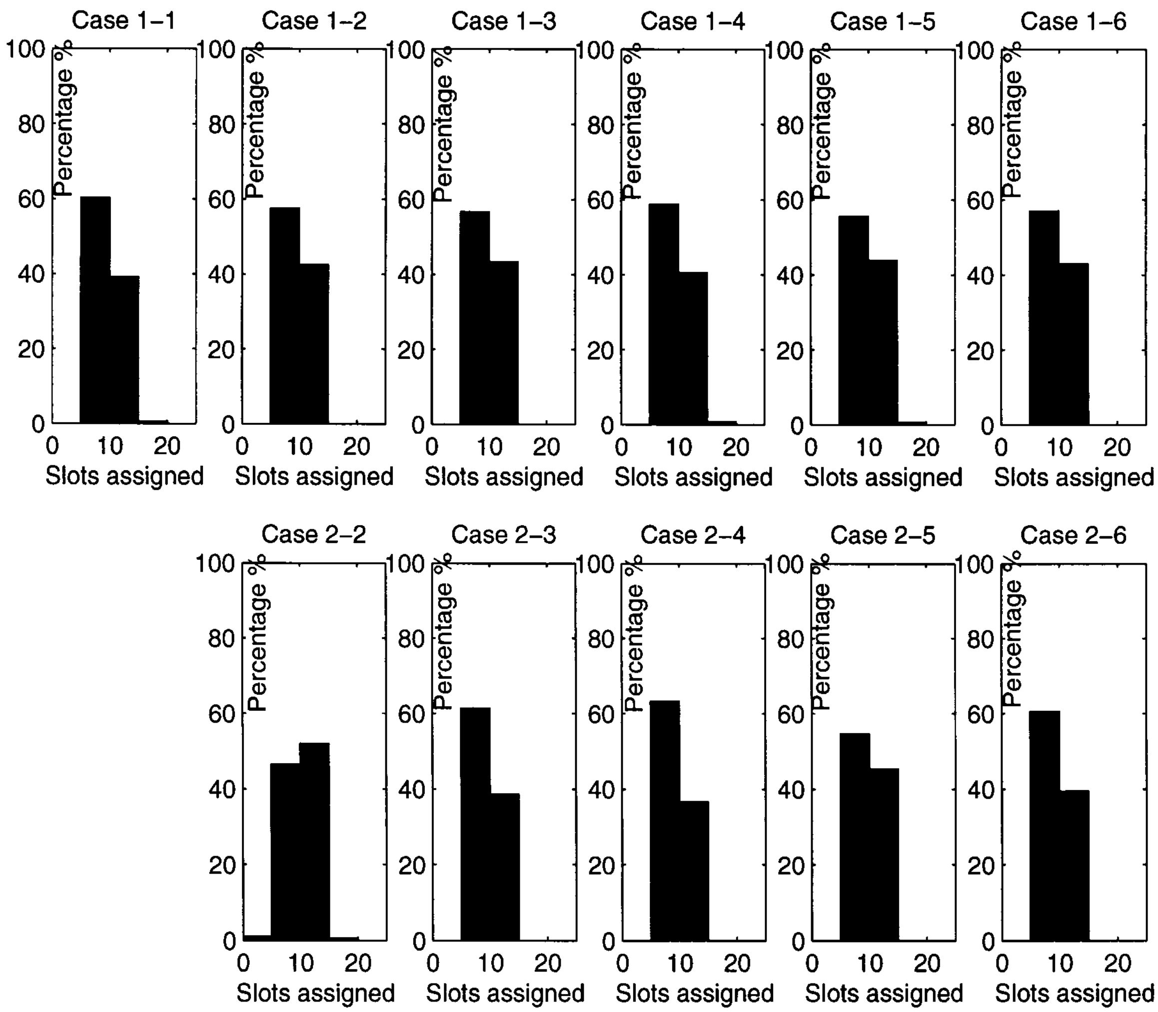

Figure 7.25: The histogram of time slots assigned-simulation with MAEs

1. The allocation approach proposed in this project can work effectively. The QoS requirements is satisfied in the simulation.

2. Compared with the conventional allocation methods such as the fixed allocation and the peak rate allocation, our allocation methods have better a performance evaluation value. In the best situation, i.e., case 2-4 with MAEs, the improvement is obvious. Compared with the peak rate allocation, the time slots utilization efficiency is improved from 6.96% to 16.74% while the band-

Table 7.3: Performances summary

	Main simulation				Simulation with BBs				Simulation with MAEs			
Case	ε %	$\bar{R}$ %	η %	Z	ε %	$\bar{R}$ %	η %	Z	ε %	$\bar{R}$ %	η %	Z
1-1	15.23±0.70	1.81±0.2	4.94±0.12	4.13	15.84±0.69	1.78±0.2	4.74±0.10	4.51	16.62±0.71	1.91±0.2	4.51±0.08	4.88
1-2	15.02±0.72	1.76±0.2	5.01±0.12	4.03	15.96±0.70	1.80±0.2	4.71±0.09	4.56	16.49±0.70	1.87±0.2	4.55±0.08	4.83
1-3	15.21±0.70	1.77±0.2	4.95±0.12	4.17	15.86±0.70	1.77±0.2	4.74±0.09	4.51	16.31±0.71	1.82±0.2	4.60±0.09	4.77
1-4	15.14±0.71	1.77±0.2	4.95±0.12	4.17	15.71±0.69	1.78±0.2	4.79±0.10	4.46	16.44±0.70	1.86±0.2	4.57±0.08	4.83
1-5	10.39±0.80	1.83±0.5	7.25±0.60	0.89	15.61±0.70	1.73±0.2	4.82±0.10	4.83	16.24±0.71	1.81±0.2	4.63±0.09	4.73
1-6	10.24±0.81	1.86±0.5	7.36±0.58	0.80	15.02±0.72	1.69±0.3	5.02±0.10	4.01	16.40±0.70	1.85±0.2	4.58±0.08	4.80
2-2	12.23±0.76	1.93±0.5	6.16±0.40	1.96	16.17±0.70	1.90±0.2	4.64±0.09	4.61	16.67±0.70	1.92±0.2	4.50±0.08	4.91
2-3	15.74±0.70	1.79±0.2	4.78±0.09	4.40	16.09±0.70	1.81±0.2	4.67±0.09	4.63	16.70±0.71	1.93±0.2	4.49±0.08	4.92
2-4	16.09±0.70	1.72±0.2	4.68±0.09	4.69	16.64±0.71	1.81±0.2	4.51±0.08	4.94	***16.74±0.71***	***1.84±0.2***	**4.49±0.08**	***4.98***
2-5	15.75±0.69	1.71±0.2	4.78±0.09	4.48	16.08±0.70	1.69±0.2	4.68±0.08	4.66	16.27±0.71	1.76±0.2	4.62±0.08	4.75
2-6	16.33±0.71	1.73±0.2	4.61±0.08	4.81	15.23±0.70	1.39±0.1	4.98±0.02	4.32	16.63±0.72	1.82±0.2	4.52±0.07	4.93

width occupation ratio is decreased from 11% to 4.62%. Compared with the fixed allocation, the packet loss rate is decreased from 4.48% to 1.84%

3. In the majority cases, the fully weighted approach work better than the partially weighted one. This indicates that it is preferable to weight both the weighted assignment and the packets loss rate of the previous periods.

4. Introducing the bounds is very useful. It helps prevent huge packet losses due to a low arrival rate in previous periods. In the case 1-5, 1-6 and 2-2 of the main simulation, the rise of the packet loss rate around period 300 led to poor performances. With either the BBs or the MAEs, this did not happen any more. That is why we have a great improvement in the performance evaluation value of these three cases.

5. We found that, MAEs are better than BBs. First of all, as shown in case 2-6 with BBs, under certain situations the time slots allocation might converge which might degrade the system performance. However, this phenomenon will never happen with MAEs. In the next place, all the cases with MAEs behave better than the cases with BBs.

6. It is then clear that cases of the 2^{nd} approach with MAEs work best. However, although case 2-4 has the highest performance evaluation value, differences among these cases are not remarkable.

Chapter 8

Conclusions

The bandwidth allocation problem is a pivotal problem in network planning. This problem is called the RWA and RWTA problem in RWNs and WDM-TDM hybrid networks respectively. In the literature, the bandwidth allocation problem is usually stated as follows: Given a set of bandwidth requirement of connections in a network, route and assign the bandwidth to these connections. Most of the time, the bandwidth requirement of a single connection is considered as fixed and known in advance. It is usually measured according to the peak arrival rate. This method will inevitably cause huge bandwidth waste.

In this project, we deal with the dynamic bandwidth allocation problem in a WDM-TDM hybrid network. The allocation is updated periodically in order to follow the traffic variations. The question we try to answer is: How many time slots are needed to assign to a connection in the next period so as to guarantee its QoS requirement. The goal is to decrease the bandwidth waste while satisfying its QoS. We want to find a simple yet effective method to do the allocation instead of an accurate but time-consuming one. We followed a practical approach to evaluate the proposed algorithm. Instead of assuming a theoretical traffic model, we used a real data trace collected in the Internet.

If we regard a time slot as a bin and the packets as the weights to be packed, the bandwidth allocation problem can be considered as a bin packing problem. However, since a real data trace is employed in this project, which means we will have neither a regular packet length distribution nor a regular packet arrival pattern, the time slots allocation problem is then a stochastic bin packing problem without a specific stochastic distribution. All the discussion about the bin packing problem up to the present is based on a certain of distribution. Therefore, we need to propose our own approach to solve the time slots allocation problem in this project.

We propose a weighted assignment where we try to figure out whether or not it will fit the traffic in the next period by comparing the predicted packet loss rate in the next period with the QoS of the connection. If the answer is negative, some adjustments should then be implemented. Otherwise, the current weighted assignment is kept.

We calculate the weighted assignment and estimate the packet loss rate for the next period based on the historical information, such as the assignment and the performance in the previous periods. We considered two kinds of approaches. In the 1^{st} approach, only the packet loss rate is weighted by the previous periods and the weighted assignment equals the current one. In the 2^{nd} approach both the weighted assignment and the packet loss rate are weighted by the historical data. In these two approaches, we play with different weighting weights and different amount of previous periods that we trace back in order to find a good option.

A performance evaluation measure is defined in order to evaluate the performance of the approach that we proposed. The function aims at maximizing the bandwidth utilization ratio while minimizing the bandwidth occupation ratio and the standard deviation of the packet loss rate, in a way that respects the QoS requirements.

The main simulations and the simulations with bounds are done to verify the performance of the methods proposed in the project. In the simulations with bounds, the idea of upper and lower bounds is introduced to prevent large transients due to changes in the arrival rate. The Bollinger Bands and the Moving Average Envelopes serve as the bounds respectively in the latter simulation.

Through the simulations, we first conclude that the QoS of the connection is satisfied by using the dynamic allocation method that we proposed. Secondly, we find that generally the fully weighted approach works better than the partially weighted one. Furthermore, the bounds introduced in the simulation are confirmed to be able to eliminate the transients and to improve the system performance. The performance of the simulation with the Moving Average Envelopes are better than those of the simulation with the Bollinger Bands which are better than those of the main simulation. In addition, the system performance is much better than that of the fixed allocation or the peak rate allocation.

Finally, regarding further work, it will be interesting to test this method with other sources, as the data trace used in this project might be a particular one, as mentioned at the end of the chapter 4. In addition, the performance with multiple connections should be studied since the method has only been verified with a single connection up to now. Again, we should notice that in order to implement the 1-bin packing technique some changes might be needed, or more research should be done on this point.

Bibliography

[1] SSJ a java library for stochastic simulation user's guide. Technical report, http://www.iro.umontreal.ca/ lecuyer/ssj/.

[2] D. Banerjee and B. Mukherjee. A practical approach for routing and wavelength assignment in large wavelength-routed optical network. *IEEE Journal of Selected Areas in Communications*, 14(5):903–908, 1996.

[3] R. A. Barry, V. W. S. Chan, K. L. Hall, E. S. Kintzer, J. D. Moores, K. A. Rauschenbach, E. A. Swanson, L. E. Adams, C. R. Doerr, S. G. Finn, H. A. Haus, E. P. Ippen, W. S. Wong, , and M. Haner. All optical network consortium-ultrafast tdm networks. *IEEE Journal of Selected Areas in Communications*, 14(5), 1996.

[4] I. Chlamtac, A. Fumagalli, L. G. Kazovsky, P. Melman, W. H. Nelson, P. Poggiolini, M. Cerisola, A. N. M. M. Choudhury, T. K. Fong, R. T. Hofmeister, C. Lu, A. Mekkittikul, D. J. M. S. IX, C. Suh, , and E. W. M. Wong. Cord:contention resolution by delay lines. *IEEE Journal of Selected Areas in Communications*, 14(5):1014–1029, 1996.

[5] M. E. Corvella and A. Bestavros. Self-similarity in world wide web traffic: evidence and possible causes. *IEEE/ACM Transactions on Networking*, 5(6):835–846, 1997.

[6] E.G.Coffman, Jr., and A.L.Stolyar. Bandwidth packing. *Algorithmica*, (29):0–88, 2001.

[7] W. H. Freeman. *For all practical purposes: mathematical literacy in today's world*. New York : W. H. Freeman, 2000.

[8] D. Gamarnik. Stochastic online bin packing problem: exact conditions for stability under the best fit heuristic. *SPECIAL ISSUE: Special section on the MAMA 2001 workshop*, 29(3):30 - 31, 2001.

[9] Z. Ghassemlooy, C.Y. Cheung, G. Swift, R. Gao, J. Wang, and A. DECKER. Lecture slides for all optical time division multiplexing — an overview. Technical report, http://www.shu.ac.uk/ocr/downloads/otdm/, 2001.

[10] A. Goel and P. Indyk. Stochastic load balancing and related problems. *IEEE Foundations of computer science 40th Annual symposium*, pages 579–586, 1999.

[11] C. Guillemot, M. Renaud, P. Gambini, C. Janz, I. Andonovic, R. Bauknecht, B. Bostica, M. Burzio, F. Callegati, M. Casoni, D. Chiaroni, F. Clerot, S. L. Danielsen, F. Dorgeuille, A. Dupas, A. Franzen, P. B. Hansen, D. K. Hunter, A. Kloch, R. Krahenbuhl, B. Lavigne, A. L. Corre, C. Raffaelli, M. Schilling, J.-C. Simon, , and L. Zucchelli. Transparent optical packet switching: The european acts keops project approach. *IEEE Journal of Lightwave Technology*, 16(12):2117–2134, 1998.

[12] N. F. Huang, G. H. Liaw, and C. P. Wang. A novel all-optical transport network with time-shared wavelength channels. *IEEE Journal of Selected Areas in Communications*, 18(10):1863–1875, 2000.

[13] V. Huang and W. H. Zhuang. Optimal resource management in packet-switching tdd cdma systems. *IEEE Personal Communication*, (12):26–31, 2000.

[14] D. K. Hunter, M. H. M. Nizam, K. M. Guild, J. D. Bainbridge, M. C. CHIA, A. Tzanakaki, M. F. C. Stephens, R. V. Penty, M. J. OMahony, I. Andonovic, and I. H. White. Waspnet: A wavelength switched packet network. *IEEE Communications Magazin*, 37(3):120–129, 1999.

[15] J.Kleinberg, Y.Rabani, and E.Tardos. Allocating bandwidth for bursty connections. *IEEE Proceeding 29th ACM Symposium on Theory of Computing*, 1997.

[16] T. Mizuike, Y. Ito, D. J. Kennedy, and L. N. Nguyen. Burst scheduling algorithms for ss/tdma systems. *IEEE Transactions on Communication*, 39(4):533–539, 1991.

[17] B. Mukherjee. *Optical communication networks.* McGraw-Hill, 1997.

[18] K. Onohara, H. Sotobayashi, K. Kitayama, and W. Chujo. Photonic time slot and wavelength-grid interchange for 10gb/s packet switching. *IEEE Photonics Technology Letters*, 13(10), 2001.

[19] A. M. Sang and S. Q. Li. A predictability analysis of network traffic. *Computer Networks*, (39):329–345, 2002.

[20] D. M. Spirit, A. D. Ellis, and P. E. Barnsley. Optical time division multiplexing: Systems and neteworks. *IEEE Journal of Selected Areas in Communications*, 21(12):51–62, 1994.

[21] R. Srinivasan and A. K. Somani. A generalized framework for analyzing time-space switched optical networks. *IEEE Journal of Selected Areas in Communications*, 20(1):202–215, 2002.

[22] T. E. Stern and K. Bala. *Multiwavelength optical networks: A layered approach.* Addison- Wesley, 1999.

[23] R. A. Thompson and P. P. Giordano. An experimental photonic time slot interchanger using optical fibers as reentrant delay-line memories. *Journal of Lightwave Technology*, LT-5:154–162, 1987.

[24] A. Tripathi and K. N. Sivarajan. Resource allocation in gprs wireless networks. *IEEE International Conference on Resource Wireless Communications*, pages 388–394, 2000.

[25] V.Paxson and S.Floyd. Wide area traffic: the failure of poisson modeling. *IEEE/ACM Transactions on Networking*, 3(3):226–244, 1995.

[26] S. Yao, S. J. B. Yoo, and B. Mukherjee. All-optical packet switching for metropolitan area networks: Opportunities and challenges. *IEEE Communications Magazine*, 39(3):142–148, 2001.

[27] T. S. P. Yum, F. Tong, and K. T. Tan. An architecture for ip over wdm using time-division switching. *IEEE Journal of Lightwave Technology*, 19(5):589–595, 2001.

[28] Y. Zhang, K. Taira, H. Takagi, and S. K. Das. An efficient heuristic for routing and wavelength assignment in optical wdm networks. *Proceedings of International Conference on Communications*, 5(5):2734–2739, 2002.